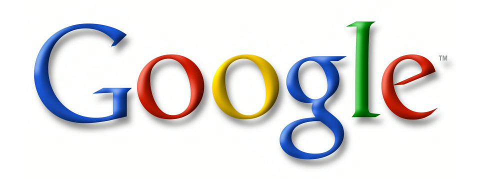
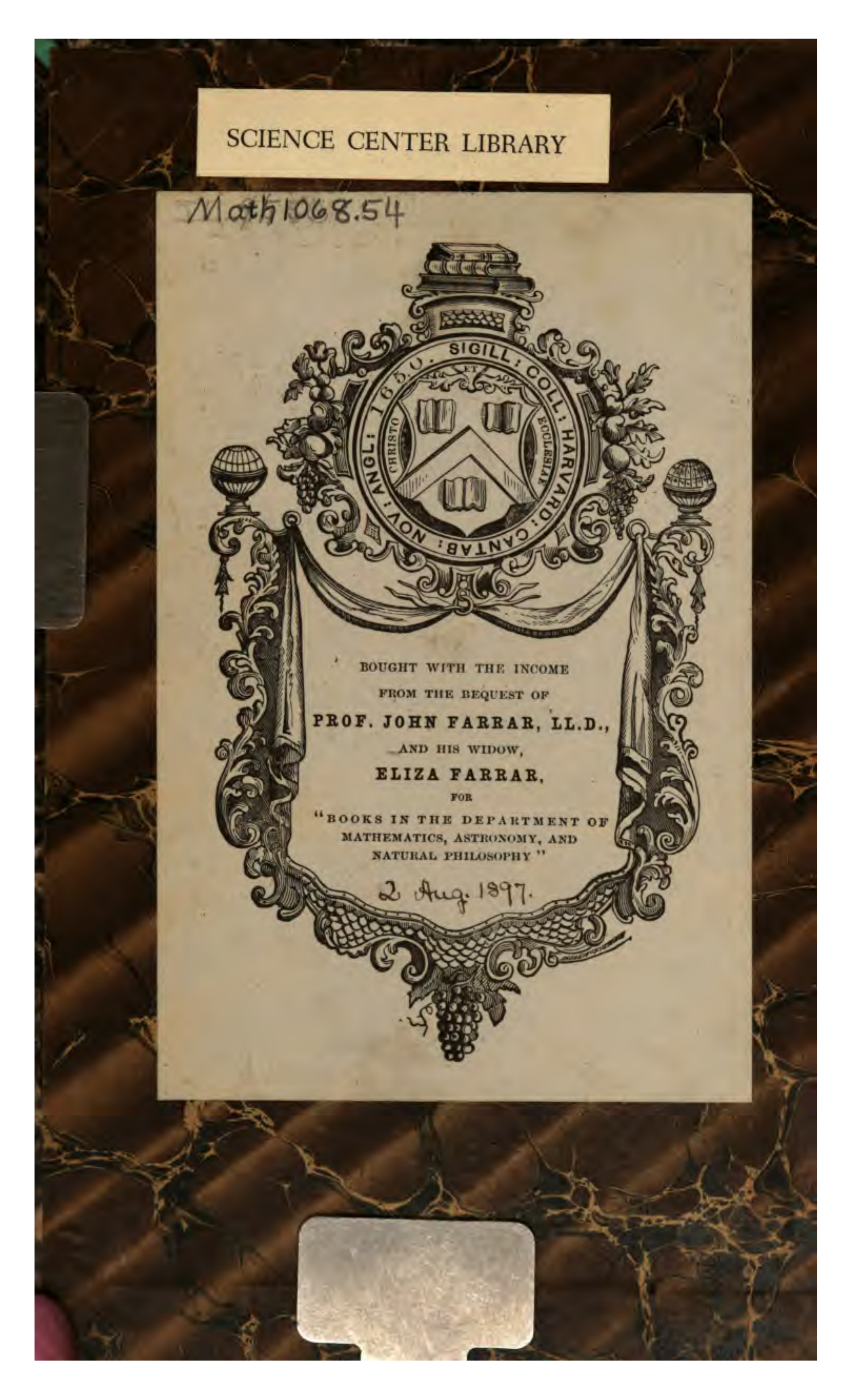
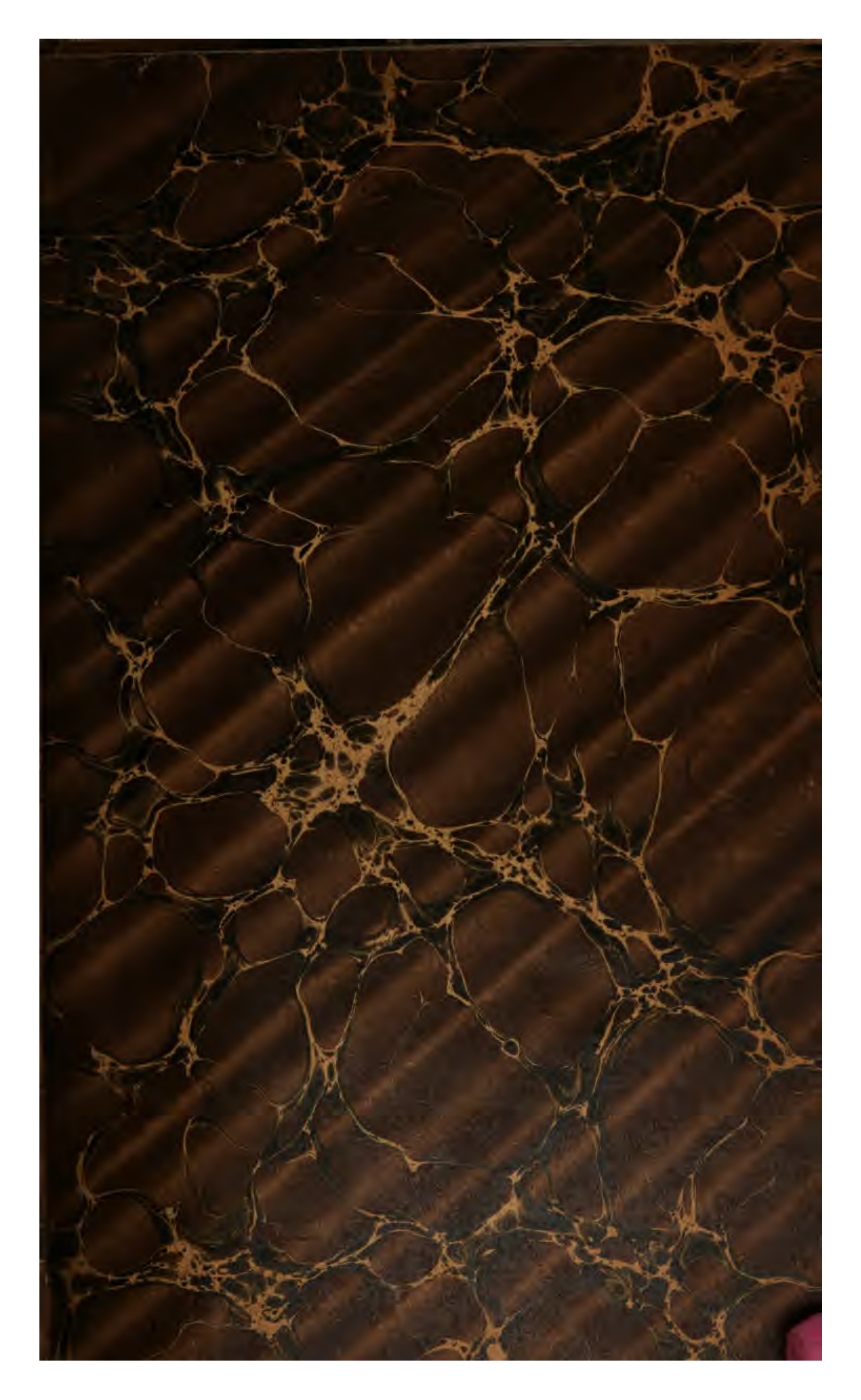

Google


SIGILL: COLL: HARVARD: CANTAB: NOV: ANGL: 1650.
CHRISTO ET ECCLESIAE



THÉORIE GÉNÉRALE

DES

APPROXIMATIONS NUMÉRIQUES.

SUIVIE D'UNE APPLICATION

A LA

RÉSOLUTION DES ÉQUATIONS NUMÉRIQUES.

OUVRAGE DU MÊME AUTEUR :

COURS COMPLÉMENTAIRE D'ANALYSE ET DE MÉCANIQUE RATIONNELLE, professé à l'École Normale. In-8°, avec planches; 1851... 7 fr.

PARIS. — IMPRIMERIE DE MALLET-BACHELIER,
rue du Jardinet, 12.

THÉORIE GÉNÉRALE

DES

APPROXIMATIONS NUMÉRIQUES,

SUIVIE D'UNE APPLICATION

A LA

RÉSOLUTION DES ÉQUATIONS NUMÉRIQUES.

A l'usage des Candidats aux Écoles spéciales du Gouvernement.

PAR M. J. VIEILLE,

Agrégé près la Faculté des Sciences de Paris, Maître de conférences à l'École Normale, Professeur de Mathématiques spéciales au Lycée impérial Louis-le-Grand.

---

seconde édition, revue, corrigée et augmentée.

PARIS,

MALLET-BACHELIER, GENDRE ET SUCCESSEUR DE BACHELIER,

Imprimeur-Libraire

DE L'ÉCOLE POLYTECHNIQUE, DU BUREAU DES LONGITUDES,

Quai des Augustins, n° 55.

1854

AVERTISSEMENT.

Cet ouvrage est surtout destiné aux jeunes gens qui se préparent aux Écoles spéciales du Gouvernement, et qui doivent faire d'abord une étude théorique approfondie des éléments des Mathématiques pour aborder utilement les applications de la science.

Le but que je me suis proposé dans la première édition de ce travail (1852), n'était pas seulement de coordonner un grand nombre de règles d'approximation qu'on pouvait trouver éparses dans les Traités d'Arithmétique. Mais il m'a semblé que dans ces Traités on s'était attaché trop exclusivement à l'*erreur absolue,* sans tenir compte de l'*erreur relative,* c'est-à-dire du rapport de la partie négligée d'une grandeur à cette grandeur même. Cependant c'est ce rapport et non l'erreur absolue qui marque le degré d'approximation obtenu.

J'ai donc cherché à rattacher à cette considération de l'erreur relative toutes les règles de l'Arithmétique appliquée aux nombres dont on n'a que des valeurs approchées. Les résultats simples auxquels je suis parvenu paraissent avoir été adoptés dans l'enseignement.

Dans cette nouvelle édition, plusieurs théories sont complétées ou éclaircies par de nombreuses applications, telles que le calcul approché du rapport de la circonférence au diamètre par la voie des séries; l'évaluation des différents degrés d'approximation fournis dans le calcul des angles par l'emploi des tangentes, sinus ou cosinus; la recherche des

triangles les plus *avantageux* à employer dans la géodésie, c'est-à-dire des triangles tels, qu'à des erreurs absolues et constantes, commises dans l'observation des angles, correspondent les erreurs relatives les plus petites possible dans les côtés calculés, etc.

La formule de Taylor est complétement démontrée pour une fonction continue quelconque d'une seule variable. Cette démonstration fort simple a été présentée à peu près sous cette forme par M. Sturm, dans ses Leçons à l'École Polytechnique. La formule générale de l'approximation, la *méthode de Newton* et *celle des parties proportionnelles* qui s'en déduisent, sont interprétées géométriquement et discutées avec soin.

Le chapitre du calcul approché des racines des équations numériques a été entièrement refondu. Les considérations géométriques auxquelles j'ai recours me permettent de fixer le sens et le degré de l'approximation fournie, soit par la méthode de Newton, soit par celle des parties proportionnelles. Pour faire ressortir les avantages qui sont propres à chacune de ces méthodes, je les applique de pair à plusieurs problèmes qui conduisent à des équations transcendantes.

Ce chapitre se termine par un exposé sommaire de la *méthode d'approximations successives par substitution*, qui consiste à négliger d'abord dans l'équation à résoudre une certaine classe de termes très-petits, auxquels on a égard ensuite en y substituant la valeur approchée de l'inconnue. Cette méthode est appliquée à plusieurs équations transcendantes, et aux équations du second et du troisième degré dans le cas où le coefficient de la plus haute puissance de l'inconnue est un nombre très-petit.

TABLE ANALYTIQUE.

THÉORIE GÉNÉRALE

DES

APPROXIMATIONS NUMÉRIQUES.

INTRODUCTION.

1. On a souvent à faire diverses opérations d'arithmétique sur des quantités dont les valeurs numériques ne sont pas exactement connues. Ainsi les données des problèmes si variés que la nature, les arts, l'industrie nous fournissent, sont toujours affectées de certaines erreurs dues à l'imperfection de nos moyens d'observation.

Ici deux questions se présentent :

1°. Si les grandeurs, commensurables ou incommensurables, qui doivent entrer dans un calcul, sont données en nombres, chacune avec un certain degré d'approximation, déterminer quel sera le degré d'approximation du résultat final ; ou, en d'autres termes, fixer une limite supérieure de l'erreur commise ;

2°. Si l'erreur admissible dans le résultat d'un calcul doit être moindre qu'une fraction assignée d'avance, déterminer le degré d'approximation avec lequel il convient d'évaluer chacune des grandeurs qui entrent dans ce calcul.

2. Quand la première question sera résolue pour les six opérations de l'arithmétique, on saura d'avance, et dans chaque cas, quel est le nombre de chiffres exacts sur lesquels on peut compter, et, par conséquent, on sera dispensé de calculer les chiffres qui expriment les unités des

ordres inférieurs. La solution de la seconde question n'est pas moins utile dans la pratique : si les éléments d'un calcul sont de la classe des incommensurables que l'on sait calculer avec une approximation illimitée, tels que les racines carrées ou cubiques, les logarithmes, le rapport π de la circonférence au diamètre, etc., il importe, pour la simplicité des opérations, de faire entrer ces nombres dans le calcul avec le moins de chiffres décimaux possible. Par exemple, soit proposé de calculer le rayon d'un cercle dont l'aire, estimée à 1 décimètre carré près, est égale à $6^{mq}45^{dq}$. En prenant le décimètre pour unité, l'expression du rayon cherché est $\sqrt{\frac{645}{\pi}}$; et il s'agira de savoir d'abord combien de chiffres exacts cette racine carrée pourra comporter; puis, quel devra être à cet effet le plus petit nombre de décimales à employer pour le diviseur π.

3. Avant d'entrer dans le détail des questions de cette nature, nous placerons quelques considérations préliminaires sur l'*erreur relative* d'un résultat numérique.

Ce qu'il importe de considérer, dans tout calcul approximatif, c'est moins l'*erreur absolue* commise que le rapport de cette erreur au résultat cherché. Nous appellerons ce rapport l'*erreur relative*. Soit α l'erreur absolue commise sur un nombre a; $\frac{\alpha}{a}$ sera l'erreur relative. C'est ce rapport seul, et non l'erreur absolue α, qui caractérise nettement le degré d'approximation obtenu. Il ne suffirait pas de savoir, par exemple, que dans la mesure d'une longueur on a fait une erreur moindre que 1 centimètre, pour en conclure que la mesure est bien prise; car, tandis que cette erreur sera regardée comme négligeable sur une longueur de 100 mètres, la même erreur sera, au contraire, très-notable, s'il s'agit d'une longueur de 1 décimètre. De même,

celui qui ferait une erreur absolue de 1 000 mètres sur la mesure de la circonférence de la terre, se tromperait beaucoup moins que celui qui ferait une erreur de $\frac{1}{100}$ de millimètre sur la mesure de la dilatation d'une barre métallique dont l'allongement total serait de 3 millimètres. Car, dans le premier cas, l'erreur relative serait $\frac{1}{40000}$, et, dans le second $\frac{1}{300}$.

Dans un calcul approximatif bien fait, l'erreur relative $\frac{\alpha}{a}$ doit être très-petite. Au reste, la petitesse de cette fraction est subordonnée au degré de précision dont on a besoin. Ainsi, dans les grandes opérations géodésiques qui ont pour objet la configuration d'une contrée fort étendue, l'ingénieur a besoin d'une base de 10 à 15 kilomètres, mesurée avec un soin extrême (à moins de $\frac{1}{100000}$ d'erreur relative par exemple), parce qu'une erreur même très-petite, commise dans la mesure de cette longueur, irait en grandissant dans le calcul des triangles du réseau dont cette base est un côté, et pourrait conduire à des résultats très-fautifs. A cet effet, on se sert de règles étalonnées que l'on pose les unes à la suite des autres sur toute la longueur de la base à mesurer. Les angles de ces triangles doivent aussi être mesurés avec la précision que l'usage du cercle répétiteur comporte $\left(\text{à moins de } \frac{1}{10} \text{ de seconde}\right)$, ce qui ne laissera subsister qu'une erreur relative moindre que $\frac{1}{100000}$ pour un angle de 15 degrés et au-dessus. Mais s'il s'agit d'un levé moins important, comme dans la topographie, on n'a plus besoin d'une aussi grande précision dans les mesures : on pourra estimer simplement les angles au graphomètre (à une minute près), et les distances à la chaîne métrique.

4. Nous supposerons, dans tout ce qui va suivre, que la valeur numérique des résultats cherchés doive être approchée *par défaut*, avec un certain nombre de chiffres entiers ou décimaux. On passera immédiatement d'une valeur par défaut à une valeur *par excès*, en ajoutant à la première la fraction qui marque la limite de l'erreur absolue. En effet, si v est une valeur de x, approchée par défaut à $\frac{1}{100}$ près, $v+\frac{1}{100}$ sera une valeur par excès de x à $\frac{1}{100}$ près.

Les deux propositions suivantes établissent la relation qui existe entre l'erreur relative d'un résultat numérique et le nombre de chiffres exacts qui lui correspond.

5. Théorème. — *Si un nombre est calculé avec* m *chiffres exacts, à partir du chiffre significatif* k *des plus hautes unités, l'erreur relative sera moindre que la fraction* $\frac{1}{k.10^{m-1}}$.

Soient a le nombre dont il s'agit, α l'erreur absolue, en sorte que $a-\alpha$ soit une valeur approchée avec m chiffres exacts; soit u l'unité de l'ordre du $m^{ième}$ chiffre à partir de la gauche; on a évidemment

$$\alpha < 1.u$$

et

$$a > k.10^{m-1}.u,$$

le signe $>$ n'excluant pas l'égalité; d'où

$$\frac{\alpha}{a} < \frac{1}{k.10^{m-1}}.$$

C. Q. F. D.

Corollaire. — On peut dire à fortiori que l'*erreur relative est moindre que* $\frac{1}{10^{m-1}}$. C'est cette limite qu'on adop-

tera dans la pratique, toutes les fois que le chiffre k ne sera pas connu immédiatement.

Remarques. — I. On peut, pour abréger l'écriture, supposer $u = 1$ dans la démonstration qui précède; et c'est ce que nous ferons dans la suite. Cela revient à prendre pour nouvelle unité principale l'unité de l'ordre du $m^{ième}$ chiffre, et ce changement n'influe en rien sur le rapport $\frac{\alpha}{a}$, puisque ses deux termes sont ainsi multipliés ou divisés par une même puissance de 10.

II. Nous avons supposé que les m premiers chiffres de a étaient connus exactement; mais le théorème n'exige pas précisément cette condition : en effet, l'inégalité $\alpha < 1$ exprime seulement que l'erreur absolue est moindre qu'une unité de l'ordre du $m^{ième}$ chiffre. Or ceci peut avoir lieu, lors même que le $m^{ième}$ chiffre de la valeur approchée serait fautif d'une unité. Ainsi, supposons que l'on prenne le nombre 3,158 pour valeur approchée de la racine carrée de 10; comme on a

$$\sqrt{10} = 3,16227\ldots,$$

le *troisième* chiffre 5 du nombre 3,158 est fautif d'une unité, ce qui n'empêche pas l'erreur absolue 0,00427... d'être moindre qu'une unité de l'ordre de ce chiffre, et, par suite, l'erreur relative d'être moindre que $\frac{1}{3.10^2}$.

6. *La réciproque* du théorème précédent *n'est pas vraie.*

Si l'erreur relative d'un nombre est moindre que $\frac{1}{k.10^{m-1}}$, il ne s'ensuit pas que les m premiers chiffres de la valeur approchée de ce nombre soient exacts, ni même que l'erreur absolue soit moindre qu'une unité de l'ordre du $m^{ième}$ chiffre.

En effet, comme on a, en général,

$$a > k.10^{m-1},$$

il est clair que l'inégalité

$$\frac{\alpha}{a} < \frac{1}{k.10^{m-1}}$$

peut subsister sans que l'on ait $\alpha < 1$.

Par exemple, soit $a = 398,2345$.

Si l'on prend pour valeur approchée 397, on fera une erreur absolue $\alpha = 1,2345\ldots$ plus grande que 1. L'erreur relative $\frac{1,2345}{398,2345\ldots}$ sera visiblement plus petite que $\frac{1}{3.10^2}$, et cependant le *troisième* chiffre 7 de la valeur approchée sera fautif.

Voici comment l'énoncé de la réciproque doit être formulé.

7. Théorème. — *Si l'erreur relative d'un nombre dont le premier chiffre est* k, *est moindre que* $\frac{1}{(k+1)10^{m-1}}$, *les* m *premiers chiffres de la valeur approchée de ce nombre seront exacts; ou, du moins, cette valeur ne sera pas fautive d'une unité de l'ordre du* m^{ième} *chiffre.*

En effet, par hypothèse, on a

$$\frac{\alpha}{a} < \frac{1}{(k+1)\,10^{m-1}};$$

mais on a, évidemment,

$$a < (k+1)\,10^{m-1};$$

donc

$$\alpha < 1. \qquad \text{C. Q. F. D.}$$

Corollaire. — Lorsque le chiffre k des plus hautes unités du nombre n'est pas en évidence, on peut remplacer la limite $\frac{1}{(k+1)\,10^{m-1}}$ par $\frac{1}{10^m}$; en effet, k est au plus égal à 9. L'énoncé du théorème précédent se change alors dans celui-ci : *Si l'erreur relative d'un nombre est moindre que*

$\frac{1}{10^m}$, *les* m *premiers chiffres de la valeur approchée de ce nombre seront exacts; ou, du moins, cette valeur ne sera pas fautive d'une unité de l'ordre du* m^{ième} *chiffre.*

8. *Remarque.* — Dans l'exemple cité plus haut, où $a = 398,2345\ldots$, si l'on prend pour valeur approchée 397,4, l'erreur relative sera moindre que $\frac{1}{4.10^2}$; ici $m=3$, et l'on peut dire, conformément au théorème, que cette valeur 397,4 n'est pas fautive d'une unité : mais on voit que le *troisième* chiffre 7 de la valeur approchée n'est pas exact. Si l'on tient à la réduire (comme d'ordinaire) à ses *trois* premiers chiffres, on ne saurait prendre 397, puisque l'erreur définitive surpasserait l'unité. On ferait ainsi deux erreurs successives et de même sens, la première résultant de ce que 397,4 n'est déjà qu'une valeur approchée par défaut, la seconde provenant de la suppression des chiffres qui suivent celui des unités. Chacune de ces erreurs est, à la vérité, moindre que 1; mais leur somme est plus grande que 1.

Cependant il y a un moyen bien simple et général de former une valeur approchée à moins d'une unité, tout en la réduisant à ses m premiers chiffres : c'est d'augmenter de 1 le dernier des chiffres conservés. Ainsi, dans le cas qui nous occupe, si l'on prend 398, les deux erreurs, au lieu de s'ajouter, se retranchent l'une de l'autre; et, comme chacune est moindre que 1, leur différence est à fortiori moindre que 1. Seulement, comme on ignore ordinairement laquelle de ces deux erreurs prédomine, on ne sait pas si le résultat définitif est approché par défaut ou par excès. Cette incertitude a peu d'importance dans la pratique; au reste, nous montrerons plus loin comment elle peut être levée.

De tout ce qui précède, on conclut la règle suivante, qui est d'un emploi fréquent :

Si l'erreur relative d'un nombre approché par défaut

est moindre que $\frac{1}{10^m}$ (*ou même, si elle est moindre que* $\frac{1}{(k+1)10^{m-1}}$, k *désignant le chiffre des plus hautes unités*), *ne conservez que les* m *premiers chiffres du nombre dont il s'agit, en ayant soin d'augmenter de* 1 *le dernier des chiffres conservés, et le résultat sera approché à moins d'une unité de l'ordre de ce dernier chiffre.*

Désormais, quand nous dirons que le résultat d'un calcul est connu avec m chiffres exacts, c'est dans le sens de cette règle qu'il faudra l'entendre.

Nous allons appliquer ces considérations générales à chacune des six opérations de l'arithmétique.

Remarque. — Dans les calculs numériques, il arrive souvent que l'on doit tenir compte de plusieurs erreurs qui se superposent. Pour que l'on puisse appliquer *à la fin du calcul* la règle précédente, il est nécessaire que les approximations aient été dirigées de manière que l'on n'ait pas de doute sur *le sens* de l'erreur finale qui affecte le résultat *non réduit* à ses m premiers chiffres.

Addition.

9. Soient a, b, c, etc., plusieurs nombres dont on a des valeurs approchées par défaut $a-\alpha$, $b-\beta$, $c-\gamma$, etc. L'erreur absolue de la somme sera

$$a+b+c+\ldots-(a-\alpha+b-\beta+c-\gamma+\ldots)$$

ou

$$\alpha+\beta+\gamma+\ldots,$$

c'est-à-dire la somme des erreurs partielles des termes a, b, c, etc. Ici le résultat $(a-\alpha+b-\beta+c-\gamma+\ldots)$ est approché par défaut. Si les erreurs partielles α, β, γ, etc., étaient, les unes par défaut, les autres par excès, les premières entreraient avec le signe + et les secondes avec le signe — dans l'expression de l'erreur finale dont la valeur

absolue serait évidemment plus petite que la somme $\alpha+\beta+\gamma+\ldots$; mais, comme on ne connaît pas d'ordinaire les valeurs mêmes des erreurs α, β, γ, mais seulement des limites que chacune ne dépasse pas, on ne pourrait plus se prononcer sur le sens de l'erreur finale.

Pour ne pas compliquer les notations, nous pouvons convenir que α, β, γ désignent respectivement les limites supérieures des erreurs commises sur a, b, c, etc., et alors nous dirons que l'erreur absolue commise sur la somme $a+b+c+\ldots$, est, *dans tous les cas,* moindre que $\alpha+\beta+\gamma$, et l'erreur relative sera moindre que

$$\frac{\alpha+\beta+\gamma+\ldots}{a+b+c+\ldots}.$$

On sait que cette fraction a une valeur moindre que la plus grande des fractions $\frac{\alpha}{a}$, $\frac{\beta}{b}$, $\frac{\gamma}{c}$. Donc l'*erreur relative d'une somme est moindre que la plus grande des erreurs relatives des termes qui la composent.*

Cette limite est simple, mais elle a l'inconvénient d'être souvent trop élevée, et, par suite, de ne pas donner une idée exacte du degré d'approximation obtenu. C'est ce qui arrivera, par exemple, si, parmi les termes a, b, c, etc., il en est un beaucoup plus petit que les autres. Ce terme influera très-peu sur la somme, et cependant son erreur relative pourra être beaucoup plus grande que celle des autres. Ainsi, supposons qu'il s'agisse d'ajouter les nombres 2,34 et 0,07, tous deux approchés à moins d'une unité du dernier ordre, c'est-à-dire à moins de $\frac{1}{100}$. L'erreur absolue de la somme 2,41 sera moindre que $\frac{2}{100}$, et l'erreur relative sera moindre que $\frac{2}{241}$ ou que $\frac{1}{100}$, tandis que la limite assignée par l'erreur relative du terme 0,07 serait $\frac{1}{7}$.

Ainsi la limite fournie par le théorème précédent ne devra être employée qu'autant que les nombres à ajouter seront du même ordre de grandeur.

Remarque. — Puisque, dans l'addition, les erreurs partielles α, β, γ, etc., s'ajoutent, il convient qu'elles soient toutes à peu près égales entre elles, c'est-à-dire que les nombres a, b, c, etc., soient tous approchés à moins d'une unité décimale du même ordre. Par exemple, si l'on a $a = 47{,}8$, à moins de $\frac{1}{10}$, et que l'on doive ajouter à ce nombre $\sqrt{3} + \sqrt{45}$, ce serait procéder peu judicieusement que de rechercher les valeurs de ces deux racines carrées à moins de $\frac{1}{10000}$;

$$\sqrt{3} = 1{,}7320, \quad \sqrt{45} = 6{,}7082.$$

La somme ainsi calculée serait 56,2402, avec une erreur moindre que 0,10002; par conséquent, on ne pourrait pas compter sur le chiffre des dixièmes, mais seulement sur celui des unités, de même que si l'on s'était contenté de calculer chaque racine avec une seule décimale.

10. Passons maintenant à la seconde question des approximations numériques : quel devra être le degré d'approximation de chaque terme d'une somme pour que celle-ci soit connue avec m chiffres exacts?

L'erreur relative de la somme doit être moindre que $\frac{1}{10^m}$ (7, Coroll.). Or on sait que cette erreur est moindre que la plus grande des erreurs relatives des termes de la somme (9); donc *il suffira que chaque terme soit évalué* à moins d'une erreur relative égale à $\frac{1}{10^m}$, c'est-à-dire *avec* $m+1$ *chiffres,* à partir du chiffre de ses plus hautes unités.

Mais, comme on l'a vu dans le paragraphe précédent, cette limite du nombre des chiffres à calculer pour chaque terme de la somme pourra être trop élevée, surtout quand ils ne seront pas du même ordre de grandeur.

C'est pourquoi, dans ce cas, au lieu de suivre la règle précédente qui assujettit les différents termes à une *même erreur relative,* il est préférable de les assujettir à une *même erreur absolue.* Soit donc α la limite inconnue que cette erreur ne doit pas dépasser, et désignons par s la somme cherchée, par p le nombre de ses termes. L'erreur relative de la somme sera moindre que $\frac{p\alpha}{s}$, et l'on posera l'inégalité

$$\frac{p\alpha}{s} < \frac{1}{10^m}, \quad \text{d'où} \quad \alpha < \frac{s}{p \cdot 10^m}.$$

Dans l'évaluation de cette limite de α, on pourra remplacer s par un nombre inférieur, que fournira immédiatement l'addition des plus hautes unités des nombres proposés.

Des deux règles que nous venons de donner pour apprécier le degré d'approximation de chaque terme d'une somme, la première est la plus facile à appliquer; elle n'exige aucun calcul. On la suivra toutes les fois que les nombres à ajouter ne différeront pas beaucoup les uns des autres.

11. *Exemples :* 1°. Soit à calculer la somme

$$\sqrt{3} + \sqrt{45} + \sqrt{167},$$

avec trois chiffres exacts.

Ici $m = 3$; en suivant la première règle, on devra donc calculer chaque racine avec *quatre* chiffres exacts; on trouve

$$\sqrt{3} = 1,732\ldots, \quad \sqrt{45} = 6,708\ldots, \quad \sqrt{167} = 12,92\ldots.$$

La somme 21,360 n'est approchée qu'à moins d'une unité de l'ordre de son troisième chiffre, c'est-à-dire à moins

de $\frac{1}{10}$. Ses deux derniers chiffres 6,0 peuvent être fautifs de plusieurs unités; c'est pourquoi on ne les conserve pas; mais, en les supprimant, il faut, conformément à la règle du n° 8, augmenter d'une unité le dernier chiffre, et l'on a

$$21,4,$$

pour la valeur demandée de la somme $\sqrt{3}+\sqrt{45}+\sqrt{167}$ réduite à trois chiffres. Elle est approchée à moins de $\frac{1}{200}$ de sa valeur (5).

La réduction que nous venons de faire a déjà été expliquée au n° 8; mais, comme elle se présentera constamment dans les calculs suivants, il ne sera pas inutile d'insister pour en bien fixer le sens. En désignant par s la somme exacte des trois radicaux, on a

$$s = 21,360 + \varepsilon, \quad \left(\varepsilon < \frac{1}{10}\right);$$

mais

$$21,360 = 21,4 - \eta, \quad \left(\eta < \frac{1}{10}\right):$$

donc

$$s = 21,4 + (\varepsilon - \eta).$$

Le signe de $\varepsilon - \eta$ n'est pas connu, mais la valeur absolue de cette différence est moindre que $\frac{1}{10}$. Par conséquent, on peut prendre 21,4 pour valeur approchée de s à moins d'une unité de l'ordre de son dernier chiffre. Seulement on ignore si 21,4 est approché par excès ou par défaut.

Si l'on tenait à lever ce doute, on n'aurait qu'à recommencer le calcul de manière à assurer une décimale exacte de plus à la somme, c'est-à-dire qu'on calculerait un chiffre de plus à chaque radical. On trouverait

$$\sqrt{3} = 1,7320, \quad \sqrt{45} = 6,7082, \quad \sqrt{167} = 12,922.$$

La somme 21,3622 serait approchée à moins de $\frac{1}{100}$, et l'on aurait

$$s = 21,3622 + \lambda, \quad \left(\lambda < \frac{1}{100}\right).$$

Or

$$21,3622 = 21,3 + \mu, \quad \left(\mu < \frac{7}{100}\right);$$

donc

$$s = 21,3 + (\lambda + \mu), \quad \left(\lambda + \mu < \frac{8}{100} < \frac{1}{10}\right):$$

21,3 est donc une valeur approchée par défaut à moins de $\frac{1}{10}$, et 21,4 l'est par excès.

Le doute sur le sens de l'erreur commise n'eût pas été levé, s'il était arrivé que le chiffre des centièmes du nombre 21,3622 fût un 9; car alors on aurait eu $\lambda + \mu < \frac{11}{100}$, d'où l'on n'aurait pu conclure que 21,3 est approché à moins de $\frac{1}{10}$. Dans ce cas, il faudrait encore recommencer le calcul avec une décimale de plus, et continuer ainsi jusqu'à ce que le chiffre de l'ordre marqué par l'approximation à laquelle on parvient, cessât d'être un 9. On ne saurait d'ailleurs trouver indéfiniment des 9; car la somme 21,399999... serait alors commensurable et exactement égale à 21,4.

Dans l'exemple qui vient d'être traité, il n'y aurait pas eu d'avantage à employer la seconde règle donnée au n° 10. Cette règle nous aurait conduit à poser

$$\alpha < \frac{s}{3.10^3}, \quad \text{ou} \quad \alpha < \frac{10}{3\ 10^3}, \quad (s \text{ surpasse } 10),$$

et, en prenant l'unité décimale de l'ordre immédiatement

inférieur,

$$\alpha < \frac{1}{1000}.$$

On aurait donc calculé chaque radical avec trois décimales, et c'est aussi ce qu'a prescrit la première règle, à l'exception de la troisième racine ($\sqrt{167}$) pour laquelle deux décimales ont suffi. On peut même remarquer que, d'après le théorème du n° 7, on aurait pu se borner à calculer $\sqrt{45}$ avec trois chiffres au lieu de quatre, et écrire $\sqrt{45} = 6,70$, attendu que le premier chiffre 6 de cette racine surpasse $k+1$ (k désignant le premier chiffre de la somme, chiffre que l'on reconnaît tout de suite ne pouvoir dépasser 2).

2°. Voici un autre exemple du cas où, les nombres à ajouter étant de grandeurs très-différentes, il faut recourir à la seconde règle.

On propose de calculer le logarithme népérien du nombre 2 avec sept décimales exactes.

On sait que ce logarithme est fourni par la série

$$l\,2 = 2\left(\frac{1}{3} + \frac{1}{3.3^3} + \frac{1}{5.3^5} + \frac{1}{7.3^7} + \ldots\right),$$

et que si l'on désigne par ε l'erreur commise en s'arrêtant au $n^{ième}$ terme de la série, on a

$$\varepsilon < \frac{1}{(2n+1)\,3^{2n-1}.4} \quad (*).$$

Dans cette question, on prévoit qu'il faudra tenir compte

(*) En général, on a pour déterminer $l(y+1)$ connaissant ly, la série

$$l(y+1) = ly + 2\left[\frac{1}{2y+1} + \frac{1}{3(2y+1)^3} + \frac{1}{5(2y+1)^5} + \ldots\right],$$

et l'erreur ε, commise en s'arrêtant au $n^{ième}$ terme, est moindre que le double produit du terme suivant $\frac{1}{(2n+1)(2y+1)^{2n+1}}$ multiplié par la somme des

de plusieurs erreurs; la première, résultant de ce qu'on ne prendra qu'un certain nombre de termes de la série; la seconde, résultant de ce qu'on négligera les chiffres décimaux d'un certain ordre dans la conversion des fractions $\frac{1}{3}$, $\frac{1}{3.3^3}$, $\frac{1}{5.3^5}$, etc., en décimales; la troisième, résultant de ce qu'on supprimera dans la somme les décimales qui suivent la septième figure.

C'est pourquoi il convient de rendre chacune de ces erreurs moindre que $\frac{1}{10^8}$, et pour cela on posera d'abord l'inégalité

$$\frac{1}{(2n+1)\,3^{2n-1}.4} < \frac{1}{10^8}$$

ou

$$4(2n+1)\,3^{2n-1} > 10^8.$$

A l'aide de quelques essais, on reconnaît que cette inégalité est satisfaite pour $n \geqq 8$.

Ainsi, en prenant les huit premiers termes de la série, on aura

$$l2 = \frac{2}{3} + \frac{2}{3.3^3} + \frac{2}{5.3^5} + \frac{2}{7.3^7} + \frac{2}{9.3^9} + \frac{2}{11.3^{11}} + \frac{2}{13.3^{13}} + \frac{2}{15.3^{15}} + \varepsilon, \quad \left(\varepsilon < \frac{1}{10^8}\right).$$

---

termes de la progression géométrique $\left[1 + \frac{1}{(2y+1)^2} + \frac{1}{(2y+1)^4} + \ldots\right]$: cette somme est $\frac{(2y+1)^2}{4y(y+1)}$; on a donc

$$\varepsilon < \frac{1}{(2n+1)(2y+1)^{2n-1}.2y(y+1)}.$$

Dans le cas de $l2$, il faut faire $y = 1$, d'où

$$\varepsilon < \frac{1}{(2n+1)\,3^{2n-1}.4}.$$

Maintenant, il faut faire en sorte que l'erreur qui affectera la somme de ces fractions, par suite de leur conversion en décimales, soit aussi moindre que $\frac{1}{10^8}$. Comme nous avons huit fractions, il suffira évidemment que chacune soit calculée avec neuf décimales.

On a d'abord

$$\frac{2}{3} = 0,666\,666\,666;$$

et, en prenant le neuvième de ce résultat,

$$\frac{2}{3^3} = 0,074\,074\,074.$$

Continuant à prendre le neuvième de chaque résultat obtenu, il vient successivement

$$\frac{2}{3^5} = 0,008\,230\,452,$$

$$\frac{2}{3^7} = 0,000\,914\,494,$$

$$\frac{2}{3^9} = 0,000\,101\,610,$$

$$\frac{2}{3^{11}} = 0,000\,011\,290,$$

$$\frac{2}{3^{13}} = 0,000\,001\,254,$$

$$\frac{2}{3^{15}} = 0,000\,000\,139.$$

Il reste à diviser les sept dernières fractions décimales res-

pectivement par 3, 5, 7, 9, 11, 13, 15, et il vient

$$\frac{2}{3} = 0,666\,666\,666$$

$$\frac{2}{3.3^3} = 0,024\,691\,358$$

$$\frac{2}{5.3^5} = 0,001\,646\,090$$

$$\frac{2}{7.3^7} = 0,000\,130\,642$$

$$\frac{2}{9.3^9} = 0,000\,011\,290$$

$$\frac{2}{11.3^{11}} = 0,000\,001\,026$$

$$\frac{2}{13.3^{13}} = 0,000\,000\,096$$

$$\frac{2}{15.3^{15}} = 0,000\,000\,009$$

Somme... $= 0,693\,147\,177$

En désignant par α l'erreur provenant des décimales négligées, nous aurons donc

$$l2 = 0,693\,147\,177 + \varepsilon + \alpha, \quad \left(\varepsilon + \alpha < \frac{2}{10^8}\right);$$

mais il faut réduire ce nombre à ses sept premiers chiffres; de là une troisième erreur $\beta < \frac{8}{10^8}$, attendu que le huitième chiffre est un 7.

Ainsi

$$l2 = 0,6931471 + \varepsilon + \alpha + \beta,$$

$$\varepsilon + \alpha + \beta < \frac{10}{10^8} \quad \text{ou} \quad < \frac{1}{10^7}.$$

Donc 0,6931471 est une valeur approchée par défaut de $l2$ à moins de $\frac{1}{10^7}$.

Si le huitième chiffre du nombre 0,693147177 avait été 8 ou 9, on aurait eu $\beta > \frac{8}{10^8}$, et, par suite, on n'aurait pas été certain que 0,6931471 approchât de $l2$ à moins de $\frac{1}{10^7}$; mais alors on aurait, conformément à la règle du nº 8, augmenté d'une unité le dernier des chiffres conservés, et le nombre résultant 0,6931472 aurait certainement eu l'approximation requise. Ici, ce nombre serait approché par excès à moins de $\frac{1}{10^7}$.

Nota. — Si l'on avait suivi dans ce calcul la première règle du nº 10, qui assujettit tous les termes de la somme à une même erreur *relative*, on aurait été conduit à calculer la fraction $\frac{2}{15.3^{15}}$ avec neuf chiffres, *à partir du chiffre significatif des plus hautes unités*, c'est-à-dire avec dix-sept décimales, puisque les huit premières sont des zéros.

On voit quelle complication cette règle eût apportée.

Soustraction.

12. Soient a et b deux nombres dont on veut calculer approximativement la différence $a - b$. Afin d'avoir une valeur approchée par défaut, nous prendrons b par excès et a par défaut. Soient $a - \alpha$ et $b + \beta$ les valeurs approchées de ces deux nombres, l'erreur absolue de la différence sera

$$a - b - [a - \alpha - (b + \beta)] \quad \text{ou} \quad \alpha + \beta;$$

les erreurs s'ajoutent, et l'erreur finale est plus petite que le double de la plus grande des deux erreurs faites sur les nombres a et b.

Si l'on avait pris les valeurs de a et b toutes deux par défaut, $a - \alpha$, $b - \beta$, l'erreur de la différence eût été $\pm (\alpha - \beta)$,

et, par conséquent, plus petite que la plus grande des deux erreurs α, β ; mais comme on n'a que des limites supérieures de α, β, et non les valeurs mêmes de ces erreurs, on ignorerait laquelle prédomine, et, par suite, on ne pourrait plus appliquer à la fin du calcul la règle de réduction du n° 8.

Il n'en est plus ici comme dans l'addition, où nous assignions une limite supérieure du résultat au moyen des erreurs relatives de ses deux termes. L'erreur relative $\frac{\alpha+\beta}{a-b}$ est évidemment plus grande que $\frac{\alpha}{a}$, et elle surpassera aussi $\frac{\beta}{b}$ toutes les fois qu'on aura $\frac{a}{b}-\frac{\alpha}{\beta}<2$. Ainsi, soient $a=120$, $b=100$, $\alpha=1$, $\beta=\frac{1}{2}$. L'erreur $\frac{\alpha+\beta}{a-b}$ ou $\frac{3}{40}$ surpasse $\frac{\alpha}{a}$ ou $\frac{1}{120}$ et $\frac{\beta}{b}$ ou $\frac{1}{200}$ (*).

Cette différence de résultats s'explique en observant que si b diffère peu de a, la différence $a-b$ pourra être très-petite, et d'un ordre très-inférieur à celui de a ou de b. Par suite, la fraction $\frac{\alpha\pm\beta}{a-b}$ pourra être beaucoup plus grande que chacune des fractions $\frac{\alpha}{a}$, $\frac{\beta}{b}$.

13. Si l'on propose de calculer la différence $a-b$ avec m chiffres exacts, quel devra être le nombre des chiffres à employer pour chacun des termes a et b ?

D'après ce qui précède, on calculera a par défaut et b par

(*) Si l'on prenait l'erreur relative de la différence, sous la forme $\frac{\alpha-\beta}{a-b}$ qui convient au cas où les deux termes a, b sont calculés par défaut, il y aurait un cas où l'erreur relative de la différence serait plus petite que la plus petite des erreurs relatives des deux termes. C'est celui où l'on aurait $\alpha>\beta$ et $\frac{\alpha}{a}<\frac{\beta}{b}$. En effet, il est aisé d'en conclure $\frac{\alpha-\beta}{a-b}<\frac{\alpha}{a}$.

excès, chacun à moins de la fraction $\frac{1}{2} \cdot \frac{a - b}{10^m}$, ou mieux, à moins de l'unité décimale immédiatement inférieure à cette fraction. (Comme a et b ne sont pas connus exactement, on remplacera, dans cette estimation, a par un nombre inférieur et b par un nombre supérieur, approchés. L'inspection du chiffre des plus hautes unités de a et de b fournira immédiatement ces deux nombres.) La soustraction faite, on saura que le résultat est approché par défaut, puis on ne conservera que les m premiers chiffres, suivant la règle du n° 8.

Maintenant, on voit aisément ce qu'il faudrait faire, si l'on avait à calculer approximativement une expression telle que

$$a + b - c + d - e,$$

composée de termes additifs et soustractifs. Pour obtenir le résultat avec m chiffres exacts, on calculera séparément la somme des termes additifs $(a + b + d)$ par défaut et la somme $(c + e)$ par excès, chacune à moins de l'unité décimale immédiatement inférieure à $\frac{1}{2} \cdot \frac{a + b + d - (c + e)}{10^m}$. Puis, l'opération s'achèvera comme dans le cas précédent.

14. Quand on emploie les Tables de Callet au calcul approché d'un produit ou d'un quotient, on est conduit à ajouter ou retrancher des logarithmes dont les erreurs absolues ont toutes une même limite supérieure, savoir : une demi-unité du septième ordre décimal, ou $\frac{5}{10^8}$; mais on ne sait pas dans quel sens est l'erreur, ce qui est une imperfection des Tables. On connaît du moins une limite supérieure de l'erreur qui affecte la somme ou la différence des logarithmes; on en déduit le nombre de chiffres exacts que comporte le retour des logarithmes aux nombres.

Soit, par exemple, à calculer par logarithmes le produit $2,9456 \times 3,4567 \times 15,234 \times 63,98 \times 46,369$.

La somme des logarithmes des cinq facteurs est

$$5,6629214,$$

et si on les suppose tous fautifs dans le même sens, de $\frac{5}{10^8}$, l'erreur absolue de la somme aura pour limite supérieure

$$\frac{25}{10^8} \text{ ou } \frac{2,5}{10^7}.$$

Ainsi, l'erreur possible sur le logarithme produit est moindre que $\left(2+\frac{1}{2}\right)$ unités du septième ordre; le logarithme qui approche le plus en moins dans la Table est 5,6629183, qui répond au nombre 46017. Ce logarithme est lui-même fautif de moins de $\frac{1}{2}$ unité du même ordre. Donc la différence 31 des deux logarithmes est fautive de moins de 3 unités. D'ailleurs la différence tabulaire 94 est approchée à moins de $\frac{1}{2}$. L'erreur correspondante du quotient $\frac{31}{94}$ aura pour limite 0,03 (comme on l'expliquera au n° 24). On ne pourra donc compter que sur le chiffre des dixièmes du quotient. La Table des parties proportionnelles fournissant pour ce quotient 0,33, on n'en conservera que le premier chiffre, et ce sera celui des unités du produit, vu la caractéristique 5. Le nombre 460173 sera la valeur approchée du produit demandé *à moins d'une unité*, c'est-à-dire *à moins de* $\frac{1}{100000}$ *de sa valeur.*

Il y a bien ici une autre cause d'erreur dont nous ne tenons pas compte; c'est *l'hypothèse de la proportionnalité* de la différence des logarithmes à la différence des nombres, proportionnalité qui n'est pas rigoureusement exacte. Mais

on verra plus loin que, pour les nombres au-dessus de 10000, l'erreur qui résulte de la règle des parties proportionnelles n'influe pas même sur les quatre premières décimales de la fraction complémentaire.

15. Comme exemple de la marche à suivre pour obtenir une approximation déterminée, dans un calcul où entrent divers termes, les uns additifs, les autres soustractifs, citons encore le *calcul approché du rapport de la circonférence au diamètre, par la série*,

$$(1) \qquad \text{arc tang}\, x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \ldots.$$

Si l'on désigne par α l'arc dont la tangente est $\frac{1}{5}$, et par β l'arc dont la tangente est $\frac{1}{239}$, on sait que l'on a

$$\frac{\pi}{4} = 4\alpha - \beta, \quad \text{ou} \quad \pi = 16\alpha - 4\beta;$$

et les valeurs de α et de β sont fournies par la série (1), dans laquelle on remplace x par $\frac{1}{5}$, puis par $\frac{1}{239}$:

$$\alpha = \frac{1}{5} - \frac{1}{3.5^3} + \frac{1}{5.5^5} - \frac{1}{7.5^7} + \frac{1}{9.5^9} - \ldots;$$

$$\beta = \frac{1}{239} - \frac{1}{3.239^3} + \frac{1}{5.239^5} - \ldots.$$

Soit proposé de calculer π avec 9 décimales. On calculera 16α par défaut à moins de $\frac{1}{10^{10}}$ et 4β par excès avec la même approximation. De cette manière, l'erreur qui en résultera pour π sera moindre que $\frac{2}{10^{10}}$, et à fortiori moindre que $\frac{1}{10^9}$.

1°. *Calcul de* 16α *par défaut à moins de* $\frac{1}{10^{10}}$. On sait

que l'erreur commise en s'arrêtant au $n^{ième}$ terme de la série (1) est moindre que le terme qui suit, ou que $\frac{x^{2n+1}}{2n+1}$. Par conséquent, pour la série qui exprime 16α, l'erreur provenant de ce qu'on s'arrêtera au $n^{ième}$ terme sera moindre que $\frac{16}{(2n+1)5^{2n+1}}$. Nous poserons donc

$$\frac{16}{(2n+1)5^{2n+1}} < \frac{1}{10^{11}},$$

afin que cette erreur, ajoutée à celles qui vont suivre, n'affecte pas la dixième décimale de 16α.

On s'assure par quelques essais que cette inégalité est satisfaite par $n \geqq 8$. Ainsi l'on prendra huit termes de cette série. Le dernier sera négatif, ce qui doit être pour que l'approximation ait lieu par défaut.

Maintenant, on calculera chacun de ces huit termes avec 12 décimales, afin que l'erreur de la somme soit moindre aussi que $\frac{1}{10^{11}}$. On aura d'abord

$$\frac{16}{5} = \frac{32}{10} = 3,2000\,0000\,0000.$$

Puis, comme $\frac{16}{5^3} = \frac{16}{5} \cdot \frac{1}{5^2} = \frac{16}{5} \cdot \frac{4}{100}$, il suffira de multiplier par 4 la fraction décimale précédente, et d'avancer de deux rangs vers la droite le produit obtenu,

$$\frac{16}{5^3} = 0,1280\,0000\,0000.$$

Le quotient $\frac{16}{5^5}$ se déduit du précédent de la même manière,

et ainsi de suite :

$$\frac{16}{5^5} = 0{,}0051\,2000\,0000,$$

$$\frac{16}{5^7} = 0{,}0002\,0480\,0000,$$

$$\frac{16}{5^9} = 0{,}0000\,0819\,2000,$$

$$\frac{16}{5^{11}} = 0{,}0000\,0032\,7680,$$

$$\frac{16}{5^{13}} = 0{,}0000\,0001\,3107,$$

$$\frac{16}{5^{15}} = 0{,}0000\,0000\,0524.$$

Il reste à diviser ces fractions, à partir de la seconde, respectivement par les nombres 3, 5, 7, 9, 11, 13 et 15, et l'on a soin de prendre par excès les quotients de deux en deux (qui sont négatifs dans la série). Il vient :

$$+\frac{16}{5} = 3{,}2000\,0000\,0000,$$

$$-\frac{16}{3.5^3} = 0{,}0426\,6666\,6667,$$

$$+\frac{16}{5.5^5} = 0{,}0010\,2400\,0000,$$

$$-\frac{16}{7.5^7} = 0{,}0000\,2925\,7143,$$

$$+\frac{16}{9.5^9} = 0{,}0000\,0091\,0222,$$

$$-\frac{16}{11.5^{11}} = 0{,}0000\,0002\,9790,$$

$$+\frac{16}{13.5^{13}} = 0{,}0000\,0000\,1008,$$

$$-\frac{16}{15.5^{15}} = 0{,}0000\,0000\,0035.$$

Ajoutant séparément les termes positifs et les termes négatifs, on a pour différence

$$3,2010\,2491\,1230 - 0,0426\,9595\,3635 = 3,1583\,2895\,7595,$$

et, par conséquent,

$$16\alpha = 3,1583\,2895\,7595 + \varepsilon, \qquad \left(\varepsilon < \frac{2}{10^{11}}\right).$$

2°. *Calcul de* 4β *par excès, à moins de* $\frac{1}{10^{10}}$.

En posant

$$\frac{4}{(2n+1)\,239^{2n+1}} < \frac{1}{10^{11}},$$

on trouve que $n \geqq 2$ satisfait; mais comme β doit être calculé *par excès*, on ne peut s'arrêter à un terme négatif de la série; il convient donc de prendre trois termes. Chacun d'eux sera calculé avec douze décimales : les deux termes positifs par excès, le négatif par défaut.

On trouve

$$\frac{4}{239} = 0,0167\,3640\,1674;$$

quant aux deux autres termes, comme ils n'auront pas plus de six chiffres significatifs (vu que la caractéristique du logarithme de $\frac{4}{(239)^3}$ est -7), on pourra s'aider des Tables ordinaires de logarithmes à sept décimales; il viendra

$$\log \frac{4}{3\,(239)^3} = \bar{8},9897450,$$

d'où

$$\frac{4}{3\,(239)^3} = 0,0000\,0009\,7666;$$

$$\log \frac{4}{5\,(239)^5} = \overline{12},01110049,$$

d'où

$$\frac{4}{5(239)^5} = 0,0000\,0000\,0002.$$

On aura donc

$$4\beta = 0,0167\,3630\,4010 - \zeta, \qquad \left(\zeta < \frac{2}{10^{11}}\right).$$

Enfin,

$$\begin{aligned}\pi &= 16\alpha - 4\beta\\ &= 3,1583\,2895\,7595 - 0,0167\,3630\,4010 + \varepsilon + \zeta\\ &= 3,1415\,9265\,3585 + \varepsilon + \zeta, \qquad \left(\varepsilon + \zeta < \frac{4}{10^{11}}\right).\end{aligned}$$

On supprime maintenant les décimales qui suivent la neuvième, et il reste

$$\pi = 3,141592653 + \eta, \qquad \left(\eta < \frac{59}{10^{11}} + \frac{4}{10^{11}} < \frac{1}{10^9}\right).$$

Ainsi la valeur approchée du rapport de la circonférence au diamètre avec neuf décimales, par défaut, est

$$3,141592653.$$

Nota. — Des trois décimales 585 que nous avons supprimées à la fin du calcul, les deux premières se trouvent encore exactes; en sorte que si l'on ne supprimait que la dernière, on aurait *de fait* le nombre π avec onze décimales. Mais nous ne pouvions répondre d'avance que ce degré d'approximation serait atteint. Si l'approximation surpasse ce que nous attendions de nos formules, cela tient à ce que nous avons constamment procédé par *à fortiori*, en posant des limites supérieures aux erreurs commises. Ainsi l'on s'assure aisément qu'en prenant trois termes de la série qui exprime 4β, l'erreur provenant des termes négligés est non-seulement moindre que $\frac{1}{10^{11}}$, mais même que

$\frac{1}{10^{16}}$; et comme d'ailleurs on n'a que trois fractions à convertir en décimales, l'erreur totale ζ qui affecte 4β est réellement moindre que $\frac{0,4}{10^{11}}$. On verrait de même que l'erreur ε commise sur 16α est moindre que $\frac{0,8}{10^{11}}$; et par suite, $\varepsilon + \zeta < \frac{1,2}{10^{11}}$. Donc, en prenant les dix premiers chiffres décimaux obtenus dans le calcul précédent, et négligeant seulement les deux derniers, on ferait une erreur définitive, moindre que $\frac{8,5 + 1,2}{10^{11}} = \frac{9,7}{10^{11}}$, ou moindre que $\frac{1}{10^{10}}$. Ainsi, l'on reconnaît bien à posteriori que

$$3,1415926535$$

est une valeur approchée par défaut de π avec dix décimales. De plus, on voit qu'en prenant onze décimales, c'est-à-dire 3,14159265358, on ne se tromperait pas de *deux* unités du dernier ordre. Mais rien n'autoriserait à adopter cette onzième décimale comme exacte.

Multiplication.

16. Théorème. — *L'erreur relative d'un produit approché par défaut est moindre que la somme des erreurs relatives des facteurs de ce produit.*

Considérons d'abord le cas de deux facteurs a, b dont les valeurs approchées sont $a - \alpha$, $b - \beta$; le produit sera approché par défaut, et l'erreur absolue sera

$$ab - (a - \alpha)(b - \beta) \quad \text{ou} \quad a\beta + b\alpha - \alpha\beta,$$

quantité plus petite que $a\beta + b\alpha$. Ainsi l'erreur absolue du produit est moindre que la somme des produits qu'on obtient en multipliant l'erreur commise sur chaque facteur par l'autre facteur.

Par suite, l'erreur relative est plus petite que

$$\frac{a\beta + b\alpha}{ab} \quad \text{ou} \quad \frac{\alpha}{a} + \frac{\beta}{b}.$$

Le théorème est donc démontré pour deux facteurs.

Prouvons maintenant que s'il est vrai pour un produit de p facteurs a, b, c, ..., l, il sera encore vrai pour un produit renfermant un facteur m de plus. Soient α, β, γ, ..., λ, μ les erreurs commises sur ces facteurs, E l'erreur relative du produit des p premiers facteurs; on a, par hypothèse, $E < \frac{\alpha}{a} + \frac{\beta}{b} + \frac{\gamma}{c} + \ldots \frac{\lambda}{l}$. Or on peut regarder le produit $abc \ldots lm$ comme composé de deux facteurs $abc \ldots l$ et m; donc l'erreur relative de ce produit sera moindre que $\frac{\mu}{m} + E$, et à fortiori moindre que $\frac{\alpha}{a} + \frac{\beta}{b} + \frac{\gamma}{c} + \ldots + \frac{\lambda}{l} + \frac{\mu}{m}$.

C. Q. F. D.

Cela posé, la proposition ayant été établie pour un produit de deux facteurs, elle sera vraie pour trois facteurs; étant vraie pour trois, elle le sera pour quatre, et ainsi de suite; donc elle est générale.

Dans l'évaluation de la somme $\frac{\alpha}{a} + \frac{\beta}{b} + \frac{\gamma}{c} + \cdots$, on substituera pour α, β, γ, etc., les limites supérieures de ces erreurs qui sont seules données, et pour a, b, c, etc., des nombres approchés par défaut.

Si quelqu'un des facteurs, tel que a, entrait dans le produit avec sa valeur exacte, on ferait $\alpha = 0$, et les conclusions précédentes subsisteraient, avec cette seule modification que, dans le cas de *deux* facteurs, l'erreur relative du produit serait précisément *égale* à celle du facteur unique dont la valeur est approchée.

17. *Si* a *et* b *sont donnés approximativement, avec* m *chiffres exacts,* quel sera le nombre des chiffres sur lesquels on pourra compter dans le produit?

On voit tout de suite que *le produit présentera toujours* m — 2 *chiffres exacts,* à partir du premier chiffre significatif à gauche. En effet, l'erreur relative de chaque facteur étant moindre que $\frac{1}{10^{m-1}}$, celle du produit sera moindre que $\frac{2}{10^{m-1}}$, et, à fortiori, moindre que $\frac{1}{10^{m-2}}$.

18. En faisant intervenir dans les erreurs relatives les premiers chiffres à gauche des deux facteurs, ainsi que le premier chiffre à gauche du produit (comme on l'a vu dans les nos 5 et 7), on peut aisément reconnaître que le produit présentera le plus souvent $m - 1$ chiffres exacts.

En effet, soient k et k' les premiers chiffres à gauche de a et de b; l'erreur relative du produit a pour limite supérieure $\frac{1}{10^{m-1}}\left(\frac{1}{k}+\frac{1}{k'}\right)$. Par conséquent, *si* k *et* k' *sont tous deux plus grands que* 1, cette erreur sera moindre que $\frac{1}{10^{m-1}}$, et dans ce cas très-étendu *on pourra compter sur les* m — 1 *premiers chiffres du produit.*

Si k *et* k' *sont tous deux égaux à* 1, l'erreur relative du produit aura pour limite supérieure $\frac{2}{10^{m-1}} = \frac{1}{5.10^{m-2}}$; mais alors le premier chiffre à gauche du produit est moindre que 4, puisque l'on a

$$a < 2.10^{m-1}, \quad b < 2.10^{m-1}, \quad \text{d'où} \quad ab < 4.10^{2m-2}.$$

Donc, eu égard au théorème du n° 7, *on pourra compter encore sur les* m — 1 *premiers chiffres du produit.*

Enfin, *si un seul des chiffres* k, k' *est égal à* 1, *la conclusion sera encore la même toutes les fois que le premier chiffre à gauche du produit ne dépassera pas* 4.

Ainsi le seul cas où l'on ne soit pas certain que le produit ait $m - 1$ chiffres exacts, est celui où un seul des chiffres k,

k' est égal à 1, tandis que le premier chiffre à gauche du produit est supérieur à 4.

Nous avons supposé que a et b étaient donnés avec le même nombre de chiffres. Dans le cas contraire, les conclusions précédentes seront encore applicables, en prenant pour m le nombre des chiffres de celui des deux facteurs qui en a le moins.

19. *Exemples.* — 1°. On propose d'évaluer l'erreur du produit

$$7850,34 \times 3,4567,$$

sachant que chaque facteur est approché par défaut, à moins d'une unité de l'ordre de son dernier chiffre.

Ici $m=5$; k et k' sont plus grands que 1. On pourra compter sur les *quatre* premiers chiffres du produit, et comme la partie entière aura cinq chiffres, on ne connaîtra ce produit qu'à une dizaine près. La multiplication, faite à la manière ordinaire, donnerait six décimales dont pas une n'est à conserver, puisque le chiffre des unités peut même être fautif. Nous reviendrons sur ce calcul après avoir exposé le procédé de la multiplication abrégée (23), qui a pour but de ne faire concourir à la formation des produits partiels que les seuls chiffres efficaces.

Le produit en question est 27136,2..., ou, d'après la règle du n° 8,

27140, à une dizaine près.

2°. Soit encore le produit

$$1850,34 \times 3,4567,$$

dont chaque facteur est approché par défaut à moins d'une unité de son dernier chiffre.

On a $m=5$, $k=1$ et $k'>1$; c'est le cas où, d'après la discussion du numéro précédent, on n'est pas certain que le

produit ait plus de $m-2$ ou de *trois* chiffres exacts. Mais si l'on prend la peine de séparer les deux erreurs relatives, on voit que celle du multiplicande est moindre que $\frac{1}{10^5}$, et celle du multiplicateur moindre que $\frac{1}{3.10^4}$. En somme, l'erreur du produit et donc moindre que $\frac{1}{10^4}\left(\frac{1}{10}+\frac{1}{3}\right)$, et à fortiori est-elle plus petite que $\frac{1}{10^4}$. Donc (7) on pourra encore compter sur *quatre* chiffres au produit, et comme la partie entière aura précisément quatre chiffres, le produit sera connu à moins d'une unité près.

3°. Considérons enfin le produit

$$9087 \times 1000,$$

dont chaque facteur est supposé approché par défaut à moins d'un unité.

On a

$$m=4, \quad k=1 \quad \text{et} \quad k'>1.$$

Ici encore on n'est pas certain à priori que le produit 9087000 ait plus de $m-2$ ou de *deux* chiffres exacts. Cependant, si l'on remarque que l'erreur absolue du produit est moindre (16) que la somme 10087 des deux facteurs, somme inférieure à 11000, on verra qu'en augmentant d'une unité le troisième chiffre 8, et négligeant les suivants, le résultat 9090000 sera une valeur approchée du produit à moins d'une dizaine de mille.

20. Le théorème démontré au n° 16 fournit immédiatement une limite supérieure de l'erreur absolue d'un produit de m facteurs a, b, c, ..., l. En la désignant par R, il est démontré qu'on a

$$\frac{R}{abc\ldots l}<\frac{\alpha}{a}+\frac{\beta}{b}+\frac{\gamma}{c}+\ldots+\frac{\lambda}{l}.$$

Chassant les dénominateurs, il vient

$$R < \alpha . bc \ldots l + \beta . ac \ldots l + \gamma . ab \ldots l + \ldots .$$

Ainsi l'*erreur absolue d'un produit de* m *facteurs approchés par défaut est moindre que la somme des produits qu'on obtient en multipliant l'erreur de chaque facteur par le produit des* m — 1 *autres facteurs.*

Dans l'évaluation de cette somme, comme il ne s'agit que d'avoir une limite supérieure, il est entendu qu'on remplacera α, β, γ, etc., par leurs limites données, et a, b, c, etc., par des nombres en excès.

Il est utile aussi, dans certains cas, notamment dans l'extraction des racines (32), d'avoir une limite *inférieure* de l'erreur absolue. Or, si l'on considère d'abord deux facteurs a, b, l'erreur absolue

$$a\beta + b\alpha - \alpha\beta$$

est évidemment plus grande que

$$(a - \alpha)\beta + (b - \beta)\alpha,$$

ou que la somme des produits de l'erreur de chaque facteur multiplié par l'autre facteur, *pris avec sa valeur approchée par défaut.* On est ainsi conduit à ce théorème :

L'erreur absolue d'un produit de m *facteurs approchés par défaut est plus grande que la somme des produits qu'on obtient en multipliant l'erreur de chaque facteur par le produit des* m — 1 *autres facteurs pris avec leurs valeurs par défaut.*

En supposant cette proposition vraie pour un produit de p facteurs a, b, c, ..., l, on prouvera, comme on l'a fait (16) pour l'erreur relative, qu'elle est encore vraie pour un produit renfermant un facteur m de plus.

21. Passons à la seconde question des approximations. Il s'agit de savoir combien de chiffres il faut employer dans

chacun des facteurs a, b, c, etc., pour que l'on puisse compter sur les m premiers chiffres du produit.

Il suffira (7) que l'erreur relative du produit effectué soit moindre que $\frac{1}{10^m}$ ou moindre que $\frac{1}{(k+1)10^{m-1}}$ (k étant le chiffre des plus hautes unités du produit). Et, pour cela, la somme des erreurs relatives des facteurs devra être, au plus, égale à cette même fraction.

L'erreur relative de chaque facteur reste donc indéterminée, pourvu que la somme ne dépasse pas cette limite. Soit p le nombre des facteurs. Ce qu'il y a de plus simple à faire, c'est de rendre chaque erreur relative inférieure à la $p^{ième}$ partie de la limite $\frac{1}{10^m}$: on posera donc

$$\left.\begin{array}{l}\frac{\alpha}{a}\\ \frac{\beta}{b}\\ \frac{\gamma}{c}\\ \cdot\\ \cdot\end{array}\right\} < \frac{1}{p \cdot 10^m},$$

ou mieux (lorsque le chiffre k sera en évidence)

$$< \frac{1}{p(k+1)10^{m-1}}.$$

Ainsi l'erreur absolue de chaque facteur sera proportionnelle (ou à peu près) à ce facteur; et, selon la grandeur de p, nous saurons combien de chiffres il faudra employer pour chaque facteur. Dans la pratique, p ne dépassera pas 4 ou 5; autrement, on devrait procéder par logarithmes. Supposons donc $p < 10$; nous pourrons formuler la règle pratique suivante:

Il suffira de m + 1 *chiffres pour tout facteur dont le premier chiffre sera égal ou supérieur à* p [ou à $\frac{p(k+1)}{10}$, si l'on peut faire usage de k. Il suffirait même de m chiffres

pour un facteur dont le premier chiffre serait égal ou supérieur à $p(k+1)$].

Dans le cas contraire, on emploiera m + 2 *chiffres.*

S'il s'agit de deux facteurs a, b, dont l'un b soit pris avec sa valeur exacte, on doit faire $p=1$, c'est-à-dire satisfaire à l'inégalité $\frac{\alpha}{a} < \frac{1}{(k+1)\,10^{m-1}}$. Ainsi, $m+1$ *chiffres suffiront toujours pour* a, *lorsque* b *sera exact.*

Dans les applications, on emploiera avec avantage le procédé de la multiplication abrégée que nous ferons bientôt connaître.

22. L'exemple suivant montrera les précautions à prendre pour éviter des calculs trop longs.

On propose de calculer avec trois chiffres exacts (ou bien à moins de $\frac{1}{100}$ de sa valeur) le produit

$$(\sqrt{3}+2)(\sqrt{45}-1)(\sqrt{8}+1).$$

On a $p=3$, $m=3$, et le premier chiffre de chaque facteur est égal ou supérieur à p. Donc (21) il suffira de $m+1$ ou quatre chiffres pour chacun.

Il vient

$$\sqrt{3}+2=3,732,$$
$$\sqrt{45}-1=5,708,$$
$$\sqrt{8}+1=3,828;$$

et si l'on désigne par P le produit exact, dont la partie entière aura évidemment deux chiffres, on aura

$$(1)\quad \begin{cases} P = 3,732.5,708.3,828+\varepsilon & \left(\varepsilon<\frac{1}{10}\right) \\ = 81,545\ldots+\varepsilon & \\ = 81,6 \pm \eta & \left(\eta<\frac{1}{10}\right). \end{cases}$$

La valeur du produit $(\sqrt{3}+2)(\sqrt{45}-1)(\sqrt{8}+1)$, à moins de $\frac{1}{100}$ de sa valeur, est donc 81,6.

Nota. — En effectuant les deux multiplications successives à la manière ordinaire, le produit 81,545... *présenterait neuf* décimales, dont *une seule* serait à conserver. Afin d'abréger ces calculs, la première question qui se présente, c'est de savoir combien de chiffres il est utile de conserver dans le premier produit, celui de 3,732 par 5,708. Or, si nous considérons le produit $P-\varepsilon$, ou

$$(3,732.5,708)\,3,828,$$

abstraction faite des radicaux dont il provient, nous n'avons besoin de connaître sa valeur qu'avec *trois* chiffres exacts; et, de plus, il faut qu'en la réduisant à ces trois chiffres seulement, nous soyons certains que cette valeur reste approchée *par défaut*, afin de pouvoir compenser à la fin du calcul l'erreur ε par une erreur de sens contraire (8).

C'est pourquoi il convient d'opérer, comme au n° 11, en assurant au produit $P-\varepsilon$, un chiffre de plus, en tout *quatre* chiffres. Or ce produit peut être regardé comme composé de deux facteurs dont *un seul* (3,732.5,708) sera pris approximativement. Donc il suffira (21) que ce dernier soit calculé avec $m+1$ ou cinq chiffres; et comme il aura évidemment deux chiffres à sa partie entière, c'est à la troisième décimale qu'on s'arrêtera. On trouve ainsi 21,302 pour valeur approchée du facteur (3,732.5,708), et il reste à la multiplier par 3,828, ce qui donne

$$P-\varepsilon = 81,545\ldots+\alpha \qquad \left(\alpha<\frac{1}{100}\right)$$
$$= 81,5+\beta \qquad \left(\beta<\frac{5}{100}+\frac{1}{100}<\frac{1}{10}\right)$$
$$= 81,6\pm\eta \qquad \left(\eta<\frac{1}{10}\right).$$

Procédé de la multiplication abrégée.

23. Dans les multiplications qu'on vient de faire, le procédé de *la multiplication abrégée* n'aurait pas d'avantage notable sur la règle ordinaire, parce que les facteurs ne sont pas composés d'un grand nombre de chiffres. C'est pourquoi, pour faire mieux ressortir la simplification attachée à ce procédé, nous préférons l'exposer sur un autre exemple.

Proposons-nous de calculer le produit

$$50,76948 \times 32,98375,$$

à moins de $\frac{1}{1000}$ de sa valeur, c'est-à-dire avec quatre chiffres exacts. Comme la partie entière aura évidemment quatre chiffres, c'est à celui des unités qu'on s'arrêtera finalement; mais il est clair que, pour pouvoir répondre du chiffre des unités du produit, il faudra connaître les dixièmes de chaque produit partiel, à cause des unités de retenues que ces dixièmes peuvent fournir. Ainsi chaque produit partiel sera calculé à moins de $\frac{1}{10}$. A cet effet, on donne ordinairement au calcul la disposition suivante :

```
           50,76948
          5738 923
          ---------
          152 307
           10 152
            4 563
              400
               15
          -------
Produit = 1674,37
```

On renverse l'ordre des chiffres du multiplicateur, qui de-

vient 5738923, et on l'écrit ainsi renversé, de manière que *le chiffre 2 de ses unités* soit *placé au-dessous du chiffre* 6 *des centièmes* du multiplicande (en général, au-dessous du chiffre du multiplicande qui exprime des unités *inférieures de deux ordres* à l'unité qui marque l'approximation cherchée). Il résulte de cette disposition que chaque chiffre du multiplicateur correspond à celui du multiplicande, dont le produit par ce chiffre donne des centièmes. Cela posé, on multiplie successivement par chaque chiffre du multiplicateur la partie du multiplicande située à gauche et terminée au rang du chiffre multiplicateur. Ainsi, pour le premier produit partiel, on n'a égard qu'à la partie 50769, et l'on dit 3 fois 9 font 27, etc., ce qui donne 152307 centièmes. Pour le second produit partiel, on n'emploie que la partie 5076 qui, multipliée par 2, donne 10152 centièmes, etc. On s'arrête au dernier des chiffres du multiplicateur qui a son correspondant supérieur dans le multiplicande, c'est-à-dire au chiffre 3, lequel donne pour cinquième et dernier produit partiel, 15 centièmes. Enfin, additionnant tous ces produits, on obtient 1674,37.

Pour apprécier l'erreur commise, remarquons que chaque produit partiel est approché à moins d'un nombre de centièmes marqué par le chiffre qui a servi de multiplicateur. Ainsi, le produit partiel 152307 centièmes est approché à moins de 3 centièmes; car la partie négligée du multiplicande est moindre qu'une unité de l'ordre du dernier chiffre employé 9; et cette unité, multipliée par le chiffre 3 du multiplicateur, donnerait 3 centièmes. D'après cela, l'erreur commise sur la somme des cinq produits partiels est moindre que

$$0{,}01\,(3+2+9+8+3);$$

mais ce n'est pas encore la limite supérieure de l'erreur totale, parce qu'il faut tenir compte de la partie 75, négligée au multiplicateur. Or cette partie est moindre qu'une unité de l'ordre du dernier chiffre employé 3; et cette unité, multipliant le 5 du multiplicande, donnerait un produit égal à 5 centièmes, avec une erreur moindre que 1 centième. Donc l'erreur résultant de ce qu'on a négligé les chiffres du multiplicateur qui n'ont pas de correspondant supérieur dans le multiplicande, est moindre que $(5+1)$ centièmes.

Enfin, l'erreur totale qui peut affecter le produit 1674,37 a pour limite supérieure,

$$0{,}01\,(3+2+9+8+3+5+1) \quad \text{ou} \quad 0{,}31.$$

Maintenant, si l'on réduit ce produit à ses quatre premiers chiffres, on fera une nouvelle erreur de 0,37 qui, ajoutée à la limite précédente, donnera $0{,}68 < 1$. Donc 1674 est la valeur par défaut du produit cherché, à moins d'une unité.

Dans le cas où l'erreur provenant de la réduction du produit à ses quatre premiers chiffres, ajoutée à la limite d'erreur précédemment obtenue, aurait donné une somme plus grande que 1, on aurait pris 1675, conformément à la règle du n° 8. Seulement, le sens de l'approximation serait resté incertain. Si l'on tenait à lever ce doute, il aurait fallu calculer le produit avec un chiffre exact de plus, c'est-à-dire avec cinq chiffres.

Remarque. — Si le multiplicande ne s'était pas prolongé vers la droite, au delà du chiffre 9 auquel correspond le dernier chiffre à droite du multiplicateur, en sorte que la partie 48 n'eût pas existé, le premier produit partiel 152307 n'eût pas été fautif; et, par suite, dans l'expression ci-dessus de la limite d'erreur, le chiffre 3 qui a servi de multi-

plicateur à ce produit partiel aurait disparu de la somme $(3+2+9+8+\ldots)$, et la limite d'erreur se serait réduite à $0,01\,(2+9+\ldots+5+1)$.

Si l'on supposait que le multiplicande perdît encore un chiffre vers la droite et se réduisît à 50,76, on le compléterait on écrivant un zéro à la droite du 6, afin que le chiffre 3 du multiplicateur eût son correspondant. Mais alors on voit que les deux premiers produits partiels cesseraient d'être fautifs, et, par suite, leurs multiplicateurs 3 et 2 cesseraient de figurer dans la limite de l'erreur qui se réduirait à $0,01\,(9+8+3+5+1)$; et ainsi de suite.

De même, si le multiplicateur ne s'était pas prolongé vers la gauche, au delà du chiffre 3 auquel correspond le premier chiffre du multiplicande, en sorte que la partie 75 n'eût pas existé, on aurait dû supprimer $(5+1)$ de la somme précédente, ce qui aurait réduit d'autant de centièmes la limite de l'erreur, etc.

De ce qui précède, on conclut cette règle générale pour apprécier le degré d'erreur que comporte le procédé de la multiplication abrégée :

On fait la somme des chiffres du multiplicateur renversé, à partir du premier de ceux des chiffres à droite dont le correspondant multiplicande est suivi d'un ou plusieurs chiffres significatifs, jusqu'au dernier de ceux qui ont un correspondant multiplicande ; et, si le multiplicateur se prolonge à gauche du multiplicande, on ajoute à cette somme le premier chiffre à gauche du multiplicande, augmenté de 1. — Puis, on multiplie le tout par l'unité inférieure de deux ordres à celle qui marque l'approximation demandée. Le produit qui en résulte est une limite supérieure de l'erreur commise.

Tant que le nombre des produits partiels ne dépassera pas 10, la somme des chiffres du multiplicateur et du pre-

mier chiffre du multiplicande, plus 1, sera au plus égale à $(9 \times 10 + 9 + 1)$ ou 100. Par conséquent, l'erreur commise sur le produit total sera moindre qu'une unité de l'ordre qui marque l'approximation demandée.

Si le nombre des produits partiels dépassait 10, l'approximation pourrait n'être plus suffisante. Il faudrait alors avancer d'un rang vers la droite le multiplicateur, de sorte que chacun des chiffres de ce facteur fût placé sous celui du multiplicande dont le produit par ce chiffre donne des unités inférieures de *trois ordres* à l'unité qui marque l'approximation demandée.

Au contraire, il pourra arriver qu'il suffise de placer le multiplicateur renversé de manière que le chiffre de ses unités soit sous le chiffre du multiplicande qui exprime des unités inférieures *d'un ordre* seulement à l'unité qui marque l'approximation demandée. Il en sera ainsi, quand la somme des chiffres à employer comme facteurs dans le multiplicateur ainsi disposé, augmentée du premier chiffre à gauche du multiplicande, sera moindre que 10. En effet, soit u l'unité de l'ordre qui marque l'approximation qu'on doit atteindre ; l'erreur commise sur le produit sera moindre que $11\frac{u}{10}$; et comme le dernier chiffre à droite qui est à négliger dans le produit, est au plus égal à 9, sa valeur relative sera au plus $9\frac{u}{10}$. Somme toute, l'erreur définitive sera moindre que $20\frac{u}{10}$ ou $2u$. Donc, en augmentant de 1 le dernier des chiffres conservés, on aura le produit à moins de u, comme il était demandé. Le sens de l'approximation restera incertain.

Il est entendu que si le multiplicande n'avait pas assez de chiffres vers la droite pour que les chiffres du multiplicateur pussent être placés selon les règles précédentes,

on compléterait le multiplicande en écrivant des zéros à sa droite.

24. Comme il importe de se familiariser avec les règles qui viennent d'être exposées, nous placerons ici d'autres exercices.

1°. Reprenons d'abord l'exemple du n° **19**. On disposera le calcul ainsi :

$$\begin{array}{r}
1850,34 \\
765\ 43 \\
\hline
5551\ 02 \\
740\ 12 \\
92\ 50 \\
11\ 10 \\
1\ 26 \\
\hline
6396,00
\end{array}$$

En considérant les deux facteurs 1850,34 et 3,4567 comme exacts, 6396 est une valeur approchée de leur produit à moins de 0,01 $(4+5+6+7)$ ou 0,22 ; mais comme ces facteurs ne sont eux-mêmes qu'approchés à moins d'une unité de l'ordre du dernier chiffre décimal, nous savons (**19**) que leur produit est déjà affecté d'une erreur qui a pour limite supérieure l'unité. Pour compenser ces deux erreurs (**8**), on augmentera d'une unité le quatrième chiffre, et 6397 satisfera à la question.

Il est à propos de remarquer que si l'on n'avait pas d'avance apprécié l'erreur qui provient de ce que les deux facteurs ne sont pas exacts, la règle donnée au n° **23** aurait pu fournir la limite de l'erreur totale. En effet, le multiplicande n'étant pas fautif d'un centième, le premier produit partiel 555102 est approché à moins de $0,01 \times 3$, et, par conséquent, on aura égard à l'approximation du multiplicande en rétablissant dans la limite de l'erreur le facteur 3, comme si le dernier chiffre 4 du multiplicande était

suivi d'autres chiffres. De même, pour tenir compte de l'approximation du multiplicateur, il suffit de remarquer que ce facteur est en défaut de moins d'une unité de l'ordre de son dernier chiffre 7, et que cette unité donnerait moins de 0,01 (18 + 1) dans le produit. Donc enfin la limite de l'erreur *totale* est

$$0{,}01\,(3+4+5+6+7+18+1), \quad \text{ou} \quad 0{,}44<1,$$

et, par conséquent, nous savons maintenant que 6396 *est une valeur approchée par défaut*, à moins d'une unité.

2°. On propose de calculer l'aire d'un cercle dont le rayon est $4^{\text{m}},72$, sachant que la valeur de ce rayon est approchée à moins de 1 centimètre.

L'aire cherchée est un produit dont l'un des facteurs est le carré du rayon, et l'autre le nombre π. Or l'erreur relative du rayon $4^{\text{m}},72$ étant par hypothèse moindre que $\frac{1}{400}$, l'erreur relative de son carré sera moindre que $\frac{2}{400}$, ou $\frac{1}{200}$, et cette dernière fraction sera nécessairement une partie de la limite de l'erreur finale. Ainsi, *quelle que soit l'approximation du facteur* π, on ne saurait répondre que l'erreur relative de l'aire sera moindre que $\frac{1}{200}$. On peut seulement se proposer de la rendre inférieure à $\frac{1}{100}$, et, pour cela, il suffira (16) que l'erreur relative de la valeur approchée de π soit moindre elle-même que $\frac{1}{200}$. On prendra donc $\pi = 3{,}14$, et le produit

$$(4{,}72)^2 . 3{,}14, \quad \text{ou} \quad 22{,}2784 \times 3{,}14,$$

sera une valeur approchée de l'aire du cercle avec deux chiffres exacts. Comme la partie entière aura évidemment

deux chiffres, on voit que l'aire sera connue à moins de 1 mètre carré seulement.

La question est ramenée à calculer le produit

$$22,2784 \times 3,14$$

avec deux chiffres exacts ou à moins d'une unité. D'après la règle de la multiplication abrégée, on disposera le calcul ainsi :

$$\begin{array}{r} 22,2784 \\ 4\ 13 \\ \hline 66\ 81 \\ 2\ 22 \\ 88 \\ \hline 69,91 \end{array}$$

Le nombre 69,91 est une valeur approchée par défaut du produit $(4,72)^2 . 3,14$, à moins de 0,01 $(3 + 1 + 4)$, ou de 0,08; et, par suite, 69 est une valeur du même produit, approchée par défaut à moins de 0,99 ou de 1. Donc enfin l'aire du cercle est égale à 70 mètres carrés, à 1 mètre carré près.

3°. On propose de calculer avec quatre chiffres exacts la longueur de la circonférence qui a pour rayon le côté du décagone régulier étoilé, inscrit dans le cercle de rayon 1.

Le nombre à calculer est exprimé par

$$\pi(\sqrt{5} + 1);$$

puisque l'on veut quatre chiffres exacts à ce produit, on prendra (21) chaque facteur avec cinq chiffres, c'est-à-dire

$$\pi = 3,1415, \quad \sqrt{5} + 1 = 3,2360.$$

Le produit ayant deux chiffres à sa partie entière, c'est au chiffre des centièmes qu'on devra s'arrêter. La multiplication abrégée fournit le nombre 10,1655 approché à moins

de 0,0011, et, en négligeant les deux derniers chiffres, on a 10,16, qui est approché à moins de 0,0066, et à fortiori à moins de 0,01. Enfin, on a

$$\pi(\sqrt{5}+1) = 10,17 \pm \varepsilon, \quad \left(\varepsilon < \frac{1}{100}\right).$$

4°. Calculer, à moins de 1 centimètre près, la longueur d'une circonférence dont le rayon est la diagonale du carré qui a 1 mètre de côté.

La longueur cherchée est exprimée par $2\pi\sqrt{2}$ ou $\pi\sqrt{8}$, et, comme ce produit sera moindre que 10, la question revient à le calculer avec trois chiffres exacts. Pour cela, on prendra chaque facteur avec quatre chiffres; soient donc

$$\pi = 3,141, \quad \sqrt{8} = 2,828.$$

Le produit calculé par le procédé abrégé est 8,8824, à moins de 0,001; et en le réduisant à 8,88, on a une valeur par défaut approchée à moins de 0,01. Donc enfin, la longueur de la circonférence cherchée est $8^{m},89$, à moins de 1 centimètre près.

Division.

25. Théorème. — *L'erreur relative d'un quotient approché par défaut est moindre que la somme des erreurs relatives de ses deux termes.*

Soient $a - \alpha$, $b + \beta$, les valeurs approchées des deux termes du quotient $\frac{a}{b}$, la première par défaut et la seconde par excès. Le quotient $\frac{a-\alpha}{b+\beta}$ sera approché par défaut, et son erreur absolue sera

$$\frac{a}{b} - \frac{a-\alpha}{b+\beta} \quad \text{ou} \quad \frac{a\beta + b\alpha}{b(b+\beta)},$$

quantité moindre que $\frac{a\beta + b\alpha}{b^2}$; et, en divisant par $\frac{a}{b}$, on voit que l'erreur relative sera inférieure à

$$\frac{a\beta + b\alpha}{ab} \quad \text{ou} \quad \left(\frac{\alpha}{a} + \frac{\beta}{b}\right).$$

C. Q. F. D.

Cette limite est la même que pour la multiplication (16); et l'on s'explique cette coïncidence en regardant $\frac{a}{b}$ comme le produit de a par $\frac{1}{b}$.

26. Si le dividende a était pris par excès ainsi que le diviseur, on ne saurait pas si le résultat $\frac{a+\alpha}{b+\beta}$ serait approché par défaut ou par excès; l'erreur relative aurait alors pour limite $\pm\left(\frac{\alpha}{a} - \frac{\beta}{b}\right)$, différence des erreurs relatives des deux termes. On pourrait donc, à fortiori, assigner à l'erreur la limite supérieure posée dans le théorème précédent; mais cette limite n'a plus lieu rigoureusement quand le diviseur b est pris avec une valeur par défaut (*).

Si le diviseur b est pris avec sa valeur exacte, il faudra faire $\beta = 0$, et l'erreur relative du quotient sera précisément égale à celle du dividende; si c'est le dividende qui est pris

(*) En général, soient a' et b' deux valeurs approchées par défaut de a et de b; α, β leurs approximations respectives. L'erreur commise sera, *dans tous les cas*, moindre que la différence

$$\frac{a'+\alpha}{b'} - \frac{a'}{b'+\beta}, \quad \text{ou} \quad \frac{b'\alpha + a'\beta + \alpha\beta}{b'(b'+\beta)},$$

et, si l'on divise par $\frac{a'}{b'+\beta}$, l'erreur relative sera toujours moindre que

$$\frac{\alpha}{a'} + \frac{\beta}{b'} + \frac{\alpha\beta}{a'b'}.$$

exactement, l'erreur relative du quotient sera moindre que celle du diviseur.

Un cas qui se présente fréquemment dans les calculs logarithmiques, est celui où il s'agit d'une fraction proprement dite $\frac{a}{b}$, dont les deux termes sont approchés à moins d'une demi-unité, sans qu'on sache dans quel sens. Alors on a

$$a < b, \quad \alpha < \frac{1}{2}, \quad \beta < \frac{1}{2}.$$

Si l'on suppose d'abord que a soit pris par défaut et b par excès, l'erreur absolue du quotient sera moindre que $\frac{1}{2} \cdot \frac{a+b}{b^2}$, et, à fortiori, moindre que $\frac{1}{2} \cdot \frac{2b}{b^2}$ ou $\frac{1}{b}$.

Si a est pris par excès et b par défaut, le quotient sera approché par excès, et son erreur absolue sera moindre que $\frac{a+\frac{1}{2}}{b-\frac{1}{2}} - \frac{a}{b}$ ou $\frac{1}{2} \cdot \frac{a+b}{b\left(b-\frac{1}{2}\right)}$. Or a est au plus égal à $b-1$; donc cette fraction est au plus égale à

$$\frac{1}{2} \cdot \frac{2b-1}{b\left(b-\frac{1}{2}\right)} \text{ ou } \frac{b-\frac{1}{2}}{b\left(b-\frac{1}{2}\right)} = \frac{1}{b},$$

même limite que ci-dessus.

Enfin, si a et b sont pris tous deux par défaut ou tous deux par excès, l'erreur du quotient sera évidemment plus petite que dans l'un des cas précédents.

Donc, en tout cas, on a pour limite supérieure de l'erreur commise, $\frac{1}{b}$, c'est-à-dire l'*unité divisée par le dénominateur de la fraction proposée.*

27. Cette limite trouve son application dans le calcul

approché du nombre correspondant à un logarithme donné. Si δ désigne la différence entre le logarithme donné et celui qui en approche le plus en moins, Δ la différence tabulaire, la règle des parties proportionnelles conduit à calculer la fraction $\frac{\delta}{\Delta}$, où δ et Δ sont approchés à moins d'une demi-unité de l'ordre de leur dernier chiffre; l'erreur absolue du quotient est donc moindre que $\frac{1}{\Delta}$, et, par suite, l'approximation est d'autant plus grande que Δ est plus grand. Dans les Tables de Callet, ce nombre Δ a *deux* ou *trois* chiffres au plus. C'est pourquoi, dans la conversion de $\frac{\delta}{\Delta}$ en décimales, on ne pousse pas le calcul au delà du chiffre des centièmes.

Dans la recherche de l'arc correspondant à un logarithme-sinus, ou logarithme-cosinus, ou logarithme-tangente, la règle des parties proportionnelles conduit à multiplier $10''$ par la fraction $\frac{\delta}{\Delta}$. L'erreur provenant de ce que δ et Δ ne sont approchés qu'à une demi-unité près, est donc une fraction de $10''$ moindre que $\frac{1}{\Delta}$; ou, ce qui revient au même, elle est une fraction de $1''$ moindre que $\frac{1}{\frac{\Delta}{10}}$. Par exemple, soit

$$\log \operatorname{tang} x = \bar{1},2323674.$$

Le logarithme qui approche le plus en moins est $\bar{1},2323189$, et il répond à l'arc de $(9^\circ 41' 20'')$. La différence tabulaire $\Delta = 1269$, et l'on a $\delta = 485$. Le nombre de secondes à ajouter à l'arc précédent est donc $10''.\frac{485}{1269}$, et, dans ce quotient, on pourra répondre de $\frac{1}{126}$ de seconde. En poussant la division jusqu'aux centièmes, on trouve $3'',82$; d'où

l'on conclut

$$x = 9^\circ 41' 23'',82.$$

Si le même arc avait dû être calculé par son logarithme-cosinus, le diviseur 1269 eût été remplacé par la différence tabulaire 36 relative au cosinus, et l'on n'aurait pu répondre dans le quotient que de $\frac{1}{3}$ de seconde.

Avec le logarithme-sinus, l'approximation eût été à peu près la même qu'avec la tangente, cependant un peu moindre, puisque le diviseur Δ eût été 1233.

Si l'on désigne par Δ, Δ', Δ'' les différences tabulaires correspondantes aux trois lignes trigonométriques tangente, sinus, cosinus d'un même arc, on a toujours

$$\Delta = \Delta' + \Delta'';$$

par conséquent, l'approximation fournie par la tangente est toujours supérieure à celle que donnent les deux autres lignes trigonométriques.

28. *Si a* et *b sont donnés approximativement chacun avec m chiffres*, quel sera le nombre des chiffres sur lesquels on pourra compter dans le quotient?

L'erreur relative du quotient ayant même limite supérieure que celle du produit, on voit tout de suite, comme au n° 17, que l'*on pourra toujours compter sur les* m — 2 *premiers chiffres du quotient*.

Puis, en faisant intervenir les premiers chiffres à gauche k et k' de a et de b, on reconnaît, comme au n° 18, que si k et k' *sont tous deux plus grands que* 1, *on pourra compter sur* m — 1 *chiffres*.

Mais si k et k' sont tous deux égaux à 1, on ne peut plus dire que le premier chiffre à gauche du quotient soit, comme celui du produit, moindre que 4. Par exemple,

$$\frac{1234}{1596} = 0,773\ldots;$$

on ne peut donc pas affirmer que le quotient présentera, dans ce cas, $m-1$ chiffres. La dernière conclusion du nº **18** doit être ainsi modifiée :

Si k *et* k′ *sont égaux à* 1, *ou que l'un d'eux seulement soit égal à* 1, *on pourra compter sur* m — 1 *chiffres au quotient toutes les fois que le premier chiffre du quotient ne dépassera pas* 4.

29. Passons à la question inverse. Combien de chiffres faut-il employer dans les termes a et b, pour que l'on puisse compter sur les m premiers chiffres du quotient $\frac{a}{b}$?

Bien que la réponse à cette question résulte implicitement de la discussion précédente, il ne sera pas inutile, comme exercice, de la chercher directement.

Il suffit que l'erreur relative du quotient soit moindre que $\frac{1}{10^m}$, et, par suite, que les erreurs relatives de a et b soient moindres que $\frac{1}{2.10^m}$.

Donc, *si le premier chiffre à gauche de* a *ou de* b *est plus grand que* 1, *il suffira de calculer ce nombre avec* m + 1 *chiffres; sinon, on en prendra* m + 2.

$m+1$ chiffres suffiront toujours pour l'un des nombres a ou b, quand l'autre sera exact.

En faisant usage du premier chiffre du quotient, chiffre que l'on peut souvent avoir à l'inspection de a et de b, les limites que nous venons d'assigner s'abaissent. En effet, *si le premier chiffre du quotient ne dépasse pas* 4, il suffit, pour que l'on puisse compter sur ses m premiers chiffres, que son erreur relative soit inférieure à $\frac{1}{(4+1)10^{m-1}}$, et, par suite, que les erreurs relatives de a et b soient moindres que $\frac{1}{2(4+1)10^{m-1}}$ ou que $\frac{1}{10^m}$. Donc alors $m+1$ *chiffres*

suffiront pour a *et* b, lors même que leur premier chiffre à gauche serait 1.

En définitive, il n'y a lieu d'employer $m+2$ chiffres que lorsque le premier chiffre de a ou de b est 1, et que le premier chiffre du quotient dépasse 4 (*).

30. *Exemples.* — 1°. On propose de calculer avec dix chiffres exacts et par défaut la valeur du module $M = \frac{1}{l10}$, par lequel on doit multiplier les logarithmes népériens pour les transformer en logarithmes vulgaires.

On a déjà vu (11) un exemple du calcul approché d'un logarithme népérien avec telle approximation qu'on voudra : il s'agissait de $l2$. Le logarithme de 10 s'obtiendra en ajoutant $l2$ à $l5$ calculé par la même méthode. Mais combien faudra-t-il de décimales exactes pour chacun de ces deux logarithmes ?

Comme on tient à connaître le sens de l'approximation du module $\frac{1}{l10}$ *réduit à ses dix premiers chiffres*, on devra opérer comme s'il s'agissait de calculer ce quotient avec *onze* chiffres. Or le diviseur $l10$ est seul approximatif; donc il suffira de $m+1$ ou *douze chiffres* pour ce diviseur. (Nous disons douze chiffres et non pas douze décimales, car la partie entière de $l10$ aura un chiffre. Ainsi c'est *onze* décimales seulement qu'il faut chercher pour $l10$.)

A cet effet, il suffira d'avoir $l2$ et $l5$ chacun avec douze décimales exactes (10). On trouve

$$l2 = 0{,}69314\ 71805\ 59\ldots,$$
$$l5 = 1{,}60943\ 79124\ 34\ldots,$$

(*) Quand le premier chiffre de l'un des termes a ou b égalera ou surpassera $2(h+1)$ (h étant le premier chiffre à gauche du quotient), m chiffres suffiront pour ce terme.

d'où

$$l\,10 = 2{,}30258\;50929\;93 + \varepsilon \quad \left(\varepsilon < \frac{2}{10^{12}}\right)$$

$$= 2{,}30258\;50929\;9 \;+ \eta \quad \left(\eta < \frac{5}{10^{12}} \quad \frac{1}{10^{11}}\right).$$

Afin d'avoir le module par défaut, on augmentera de 1 le onzième chiffre de $l\,10$, en sorte qu'on adoptera pour diviseur

$$2{,}30258\;509300.$$

On aura ainsi

$$\frac{1}{l\,10} = \frac{1}{2{,}30258509300} + \alpha, \quad \left(\alpha < \frac{1}{10^{11}}\right).$$

Le quotient évalué en décimales donne

$$0{,}43429\;44819\;02\ldots.$$

Donc, en le réduisant à ses dix premiers chiffres, on aura

$$\frac{1}{l\,10} = 0{,}43429\;44819 + \zeta, \quad \left(\zeta < \frac{2}{10^{11}} < \frac{1}{10^{10}}\right).$$

Ainsi la valeur demandée du module est

$$M = 0{,}43429\;44819.$$

Nota. — Si l'on conservait les onze premiers chiffres du quotient, et qu'on augmentât de 1 le dernier, on pourrait adopter 0,43429 44819 1 pour valeur du module à moins de $\frac{1}{10^{11}}$; mais on ne saurait pas si cette valeur est approchée par défaut ou par excès.

En multipliant par le module, ainsi calculé avec dix chiffres exacts et par défaut, les logarithmes népériens de la suite naturelle des nombres calculés eux-mêmes avec dix chiffres exacts, on aura (18) les valeurs des logarithmes vulgaires, approchées à moins d'une unité du huitième ordre décimal. Puis, en supprimant les chiffres décimaux qui suivent le septième, et augmentant ce dernier de 1 toutes les fois que le huitième sera égal ou supérieur à 5, on obtiendra les logarithmes vulgaires approchés à moins de

$\frac{1}{2.10^7}$, tels qu'ils sont donnés dans les Tables de Callet. A partir du nombre 8 et pour les nombres supérieurs, les logarithmes népériens ont une partie entière égale ou supérieure à 2 ; et comme le premier chiffre à gauche du module est aussi plus grand que 1, on pourra toujours compter sur $m - 1$ ou neuf chiffres dans les produits dont il s'agit. Toutefois ceci ne veut pas dire que les logarithmes vulgaires seront ainsi connus à moins d'une unité du *neuvième ordre décimal ;* car, à partir du logarithme de 10, qui est égal à 1, les logarithmes des nombres 11, 13, 17, etc., auront une caractéristique qui comptera pour un dans le nombre des chiffres exacts que présenteront ces mêmes logarithmes.

On emploiera avec succès, dans ce genre de calcul, le procédé de la multiplication abrégée (23), lequel rendra bien compte du degré d'approximation obtenu. Soit, par exemple, à calculer le logarithme vulgaire de 7. Les deux facteurs du produit sont

$$l7 = 1,94591\ 0149,$$
$$\mathrm{M} = 0,43429\ 44819.$$

Comme on veut conserver 8 chiffres dans le produit, on écrira le module M renversé, au-dessous du multiplicande $l7$, de manière que le chiffre 4 des dixièmes corresponde au neuvième chiffre décimal de ce multiplicande :

```
1,94591 0149
9 18449 2434
------------
7 78364 0596
  58377 3042
   7783 6404
    389 1820
    175 1319
      7 7836
        7780
        1552
          19
           9
------------
0,84509 80377
```

L'erreur totale de ce produit, eu égard au degré d'approximation de chaque facteur, a pour limite supérieure (**24**)

$$\frac{1}{10^{10}}(4+3+4+2+9+4+4+\dot{8}+1+9+2),$$

ou

$$\frac{50}{10^{10}}=\frac{5}{10^{9}}.$$

On a donc

$$\log 7 = 0,84509\,80377 + \alpha \quad \left(\alpha < \frac{5}{10^9}\right)$$
$$= 0,84509\,803 + \beta \quad \left(\beta < \frac{1,3}{10^8}\right).$$

On voit par là qu'on ne peut pas répondre de l'exactitude du *huitième* chiffre 3; mais comme il n'est pas fautif de *deux* unités, on est sûr que sa vraie valeur est moindre que 5; et, par suite, en le supprimant, on aura

$$\log 7 = 0,84509\,80, \quad \text{à moins de } \frac{1}{2.10^7}.$$

2°. Dans la construction des Tables trigonométriques, on a besoin de connaître la valeur de l'arc de 10 secondes dans le cercle dont le rayon est 1, avec 13 décimales exactes. L'expression de cet arc est $\frac{\pi}{64800}$, et comme les quatre premières décimales sont des zéros, le nombre m des chiffres à calculer, *à partir du premier chiffre significatif*, est seulement $13-4$ ou 9. On prendra donc le dividende π avec $m+1$ ou *dix* chiffres; soit $\pi = 3,141592653$. Le quotient $\frac{3,141592653}{64800}$, converti en décimales, donnera

$$0,0000484813681,$$

pour valeur de l'arc de 10 secondes, à moins de $\frac{1}{10^{13}}$.

Dans un cas simple comme celui qui précède, on peut préférer la recherche directe de l'erreur absolue à l'emploi de la règle générale. Alors on dira : Soit α l'erreur à commettre sur π ; l'erreur qui en résultera pour le quotient est $\frac{\alpha}{64800}$; or elle doit être moindre que $\frac{1}{10^{13}}$: donc il faut que α satisfasse à l'inégalité $\frac{\alpha}{64800} < \frac{1}{10^{13}}$, d'où $\alpha < \frac{64800}{10^{13}}$; et, pour cela, il suffit que α soit moindre que $\frac{10^4}{10^{13}}$ ou $\frac{1}{10^9}$. On prendra donc la valeur de π avec neuf décimales, ou avec dix chiffres, y compris la partie entière, résultat conforme à celui déjà obtenu.

3°. On propose de calculer avec trois chiffres exacts le rayon d'une circonférence dont la longueur est exprimée par $\sqrt{3}$.

La valeur du rayon est $\frac{\sqrt{3}}{2\pi}$, et l'on a

$$\sqrt{3} = 1,7320\ldots, \quad 2\pi = 6,283\ldots.$$

En divisant 17 par 6, on voit tout de suite que le premier chiffre k du quotient est 2. Par conséquent (29) il suffira de prendre m ou *trois* chiffres pour le diviseur dont le premier chiffre est 6; et, pour le dividende dont le premier chiffre est 1, on prendra $m+1$ ou *quatre* chiffres. Les valeurs approchées de $\sqrt{3}$ et de 2π qui entreront dans le calcul, seront donc respectivement

$$1,732 \quad \text{et} \quad 6,29.$$

En divisant 1732 par 629, on trouve pour quotient 275..., abstraction faite de l'ordre des unités qu'il représente. Enfin, on augmentera d'une unité le troisième chiffre, et rétablissant la virgule, on aura

$$0,276$$

pour la valeur cherchée du rayon, à moins d'un millième de l'unité linéaire.

Procédé de la division abrégée.

31. Lorsque le dividende et le diviseur sont composés d'un grand nombre de chiffres, il n'est pas nécessaire, pour avoir l'approximation demandée au quotient, de les conserver tous jusqu'à la fin de l'opération. On peut alors employer avec avantage le procédé de la division abrégée que nous allons exposer.

Soit proposé de calculer le quotient

$$\text{(A)} \qquad \frac{120627295,46}{147356,785},$$

à moins de $\frac{1}{10000}$ de sa valeur, c'est-à-dire avec cinq chiffres exacts. Il suffira de $m + 2$ ou *sept* chiffres, tant au dividende qu'au diviseur (29); faisons abstraction des virgules, puisqu'il sera facile d'en tenir compte à la fin du calcul, et augmentons d'une unité le septième chiffre du diviseur, afin que le quotient soit approché par défaut. La question est ramenée à calculer les cinq premiers chiffres entiers ou décimaux du quotient

$$\text{(B)} \qquad \frac{1206272}{1473568},$$

dont les deux termes seront traités comme des nombres exacts, et nous savons que ce quotient sera approché par défaut, à moins d'une unité de l'ordre de son cinquième chiffre. Voici le tableau des opérations :

	467
12062720	1473568
2741760	818606
1268192	
89336	
920	
32	

Il n'y a rien de changé à la règle ordinaire pour les *deux premiers chiffres* du quotient. Ainsi, après avoir ajouté un zéro à la droite du dividende (vu qu'il est plus petit que le diviseur), on a divisé 12062720 par 1473568, ce qui a donné le chiffre 8 et le reste 274176 ; on a ajouté un zéro à la droite de ce reste, et une seconde division a donné le chiffre 1 avec le reste 1268192. Il y a encore trois chiffres à trouver; or ils ne sont autres que les trois premiers du quotient $\frac{1268192}{1473568}$, et l'on sait que pour les obtenir il n'est pas nécessaire de conserver sept chiffres au diviseur; les quatre premiers suffiraient (29), puisqu'ici le dividende 1268192 est considéré comme exact, et que le diviseur seul sera altéré. Mais, afin que les erreurs qui vont s'accumuler dans le cours du calcul ne puissent rendre douteux le cinquième chiffre du quotient demandé, il convient de supprimer seulement le dernier chiffre 8 du diviseur, que l'on barre d'un trait, et l'on a soin de remplacer l'avant-dernier par le chiffre supérieur d'une unité, 7, que l'on écrit au-dessus, c'est-à-dire que l'on divise 1268192 par 147357. De cette manière le quotient ne cessera pas d'être approché par défaut, et l'erreur possible sera moindre qu'une unité inférieure *de deux ordres* à l'unité du cinquième chiffre du quotient. La division de 1268192 par 147357 fournit le chiffre 8 et le reste 89336. En raisonnant sur ce reste comme sur le précédent, on est conduit à biffer le dernier chiffre du nouveau diviseur, qui se réduit à 14736, et à diviser par ce nombre le reste 89336. L'erreur qui résulte de cette nouvelle simplification est de même sens que la précédente, et a aussi pour limite supérieure $\frac{1}{100}$ de l'unité du cinquième chiffre du quotient. On trouve ainsi le chiffre 6 et le reste 920 qui, divisé par 1474, donne le cinquième chiffre 0. Il est utile, pour apprécier l'erreur qui proviendra du rejet de la fraction complémen-

taire $\frac{920}{1474}$, de calculer encore le chiffre suivant, en divisant 920 par 148 : ce chiffre est 6. Cela posé, 0,81860 est une valeur par défaut du quotient (B), approchée à moins d'une unité de l'ordre du cinquième chiffre.

En effet, si l'on désigne par u cette unité, on a successivement

$$\frac{1206272}{1473568} = 0{,}81 + \frac{1268192^{\text{centièmes}}}{1473568} = 0{,}81 + \frac{1268192^{\text{mill.}}}{147357} + \alpha_1$$

$$\left(\alpha_1 < \frac{u}{100}\right)$$

$$= 0{,}818 + \frac{89336^{\text{mill.}}}{147357} + \alpha_1 = 0{,}818 + \frac{89336^{\text{dix-mill.}}}{14736} + \alpha_2$$

$$\left(\alpha_2 < \frac{2u}{100}\right)$$

$$= 0{,}8186 + \frac{920^{\text{dix-mill.}}}{14736} + \alpha_2 = 0{,}8186 + \frac{920^{\text{cent-mill.}}}{1474} + \alpha_3$$

$$\left(\alpha_3 < \frac{3u}{100}\right)$$

$$= 0{,}81860 + \frac{920^{\text{cent-mill.}}}{1474} + \alpha_3.$$

Ce tableau met en évidence l'accumulation des erreurs qui résultent des réductions opérées sur les diviseurs successifs. Dans notre exemple, l'erreur α_3 a pour limite supérieure $\frac{3u}{100}$. En général, sur un quotient de m chiffres, l'erreur apportée par cette méthode aura pour limite supérieure $\frac{(m-2)u}{100}$: mais il faut y joindre l'erreur provenant du rejet de la fraction complémentaire. Dans l'exemple qui nous occupe, le premier des chiffres que donnerait cette fraction $\frac{920}{1474}$ serait un 6, et ce chiffre représente des unités égales à $\frac{u}{10}$; donc l'erreur totale qui affecte le quotient 0,81860

est moindre que

$$\left(\frac{7u}{10}+\frac{3u}{100}\right) \quad \text{ou} \quad \frac{73u}{100},$$

et, à fortiori, moindre que u. C. Q. F. D.

Revenons enfin au quotient (A); nous devrons, conformément à la règle du nº 8, augmenter d'une unité le dernier chiffre du quotient précédent, et rétablissant la virgule à son rang, nous aurons

818,61

pour la valeur approchée qu'il s'agissait d'obtenir.

En ayant soin, comme nous l'avons prescrit, d'augmenter d'une unité l'avant-dernier chiffre du diviseur, en même temps qu'on supprime le dernier, tous les quotients successifs sont approchés par défaut, et il n'est pas possible qu'on rencontre 10 pour quotient de l'une des divisions partielles, ce qui pourrait au contraire arriver si l'on ne prenait pas la précaution indiquée.

Règle générale. — *Lorsque le dividende et le diviseur sont composés d'un grand nombre de chiffres entiers ou décimaux, et que l'on veut seulement obtenir au quotient* m *chiffres exacts à partir du chiffre significatif des plus hautes unités, on ne conserve que les* m + 2 *premiers chiffres du dividende et du diviseur, et, en faisant abstraction des virgules, on calcule comme à l'ordinaire les deux premiers chiffres du quotient. Cela fait, on supprime le dernier chiffre du diviseur, en augmentant d'une unité l'avant-dernier, et l'on divise, par le nombre résultant, le reste qui provient du calcul des deux premiers chiffres du quotient. Cette division fournit le troisième chiffre du quotient.*

Pour avoir le suivant, on opère sur le diviseur précédent comme sur le diviseur primitif, c'est-à-dire que l'on supprime son dernier chiffre en ayant soin d'augmenter

l'avant-dernier d'une unité ; puis on divise, par le nombre résultant, le reste de la division précédente ; et ainsi de suite.

Enfin, on augmente d'une unité le m^{ième} *chiffre trouvé au quotient.*

Remarques. — I. Dans l'exemple ci-dessus, la fraction complémentaire eût encore été négligeable, lors même que sa valeur aurait atteint $\frac{97}{100}u$. En général, puisque l'erreur apportée par la division abrégée sur un quotient de m chiffres a pour limite $\frac{m-2}{100}u$ (u désignant toujours l'unité du $m^{ième}$ chiffre), *tant que la fraction complémentaire ne dépassera pas* $\left(1-\frac{m-2}{100}\right)u$, *elle sera négligeable*, sans que le quotient cesse d'avoir l'approximation requise. C'est ce qui arrivera d'ordinaire. Cependant, dans le cas contraire, la règle précédente devrait être modifiée, en ce sens que les réductions sur le diviseur ne devraient commencer qu'après avoir trouvé, par la règle ordinaire, *les trois premiers chiffres* du quotient, etc.

II. Si l'on se contentait de chercher seulement le *premier chiffre* du quotient (au lieu des deux premiers que prescrit notre règle), puis que l'on procédât immédiatement aux réductions successives du diviseur, comme elles viennent d'être indiquées, l'erreur apportée par ce mode de division abrégée, aurait pour limite supérieure $\frac{m-1}{10}u$ (u ayant le même sens que ci-dessus). Par exemple, dans le calcul qui vient d'être fait, cette limite serait $\frac{4}{10}u$; et, eu égard à la fraction complémentaire qui est moindre que $\frac{7}{10}u$, on aurait $\frac{11}{10}u$ pour limite supérieure de l'erreur totale. On ne pourrait donc plus affirmer que cette erreur est moindre

que u. Par suite, on ne connaîtrait plus le quotient (B) *par défaut*, à moins de u, ce qui est cependant nécessaire pour qu'on puisse remonter au quotient (A). Ainsi ce procédé serait souvent défectueux.

32. Nous indiquerons ici une formule d'approximation très-simple, dont l'usage est fréquent; elle s'applique au *cas où le diviseur surpasse peu l'unité,* et elle a pour effet de *remplacer ce diviseur par un multiplicateur inférieur à l'unité de la même quantité.*

Soit proposé de calculer approximativement le quotient $\frac{a}{1+x}$, a désignant un nombre quelconque et x un nombre très-petit dont la valeur exacte peut n'être pas connue. Si d'abord on néglige x, on aura une première valeur approchée par excès, a.

Pour approcher davantage, divisons a par $1+x$, en nous arrêtant au second terme du quotient; il vient

$$\frac{a}{1+x} = a - ax + \frac{ax^2}{1+x} = a(1-x) + \frac{ax^2}{1+x}.$$

Si l'on prend $a(1-x)$ pour valeur approchée du quotient, on néglige la fraction $\frac{ax^2}{1+x}$. L'erreur *relative,* qui correspond à cette seconde approximation, est donc

$$\frac{ax^2}{1+x} : \frac{a}{1+x}, \quad \text{ou, simplement,} \quad x^2.$$

Elle est d'un *ordre* de grandeur inférieur à celui de x, ce qui est la condition d'une bonne approximation.

Si, par exemple, on a $x < 0{,}031$, d'où $x^2 < 0{,}000961$, l'erreur commise en prenant $a(1-x)$ pour le quotient sera moindre que la millième partie de la valeur de ce même quotient.

Corollaire. — Si l'on considère la fraction $\frac{a}{a+b}$, dans laquelle b est supposé petit relativement à a, on aura sen-

siblement

$$\frac{a}{a+b} = 1 - \frac{b}{a}.$$

En effet, $\frac{a}{a+b} = \frac{1}{1+\frac{b}{a}}$; on peut donc appliquer à ce quotient la formule ci-dessus. L'erreur relative sera $\frac{b^2}{a^2}$.

33. En physique, lorsqu'on veut réduire à la température zéro le volume d'un gaz mesuré à une autre température et conservant la même pression, on a, d'après la loi de Gay-Lussac, la formule

$$V_0 = \frac{V}{1+\alpha t};$$

V_0 et V désignent les volumes du gaz aux deux températures 0° et t°, et α le coefficient de dilatation. Or α est une très-petite fraction moindre que 0,0037. Ainsi, tant que le nombre t sera peu élevé, on pourra remplacer la formule précédente par celle-ci, dont le calcul est plus simple,

$$V_0 = V(1 - \alpha t),$$

et l'erreur sera une fraction du volume cherché marquée par $(\alpha t)^2$, ou moindre que $0{,}00001369\, t^2$.

Dans la règle d'escompte, la valeur actuelle d'un billet dont le montant est a, payable au bout du temps t, le taux d'intérêt étant de r pour 1 franc, est donnée par la formule

$$a(1 - rt), \quad \text{escompte commercial.}$$

Si l'on adoptait le mode d'escompte plus rationnel dans lequel on prendrait pour valeur actuelle du billet la somme qui placée pendant le temps t reproduirait a, cette valeur actuelle serait exprimée par

$$\frac{a}{1+rt}.$$

Il résulte de ce qui précède, que :

Pour des valeurs suffisamment petites de r et de t, ces deux formules s'accorderont à donner sensiblement la même fraction de a. Si, par exemple, $r = 0,05$ et que le nombre d'années t ne dépasse pas 2, la différence des nombres fournis par les deux modes d'escompte sera moindre que la centième partie de la valeur du plus grand de ces nombres.

Élevation aux puissances.

34. Quand on doit calculer approximativement une puissance d'un nombre de plusieurs chiffres, ce qu'il y a de plus simple ordinairement, c'est de recourir aux logarithmes ; on est ainsi ramené au cas de la multiplication déjà traité. Cependant il est utile de savoir assigner une limite de l'erreur commise sur a^n, lorsqu'au lieu de a on emploie la valeur approchée $a - \alpha$. Cette limite se déduit immédiatement du théorème du n° **20**, relatif au produit de plusieurs facteurs ; il suffit de supposer que les n facteurs a, b, c, etc., deviennent égaux entre eux, et l'on voit que l'erreur absolue de a^n sera moindre que $na^{n-1}.\alpha$, et l'erreur relative moindre que $n\frac{\alpha}{a}$. Ainsi l'on a ce théorème :

L'erreur relative d'une puissance d'un nombre est moindre que le produit de l'erreur relative de ce nombre par le degré de la puissance.

35. Les seules applications usuelles de ces limites concernent les carrés et les cubes.

Exemples. — 1°. Sachant qu'une valeur approchée de la toise en mètre est $1^m,9490$ à moins d'une unité du dernier ordre décimal, on propose d'évaluer la toise carrée en mètres carrés.

On a $a < 2$, $\alpha < \frac{1}{10^4}$; donc l'erreur absolue sur la valeur

de la toise carrée sera moindre que $\frac{2.2}{10^4} = \frac{4}{10^4}$: on pourra donc compter sur les trois premières décimales du carré de 1,9490,

$$(1,9490)^2 = 3,798601 ;$$

donc

$$1^{tq} = 3^{mq},799 \pm \varepsilon, \quad (\varepsilon < 0,001).$$

Si l'on veut ne conserver que les décimètres carrés, on augmentera d'une unité le chiffre des centièmes, et il viendra

$$1^{tq} = 3^{mq},80,$$

valeur approchée par excès à moins d'un demi-décimètre carré près.

Ici la considération de l'erreur relative n'aurait pas eu d'avantage sur celle de l'erreur absolue.

2°. Les mêmes choses étant posées que dans le problème précédent, on demande la valeur du pied carré en centimètres carrés.

D'après le calcul ci-dessus, on peut poser

$$1^{tq} = 3^{mq},7986 + \eta, \quad \left(\eta < \frac{5}{10^4}\right).$$

En prenant la trente-sixième partie de ce résultat, on aura la valeur du pied carré,

$$1^{pq} = 0^{mq},1055 + \zeta, \quad \left(\zeta < \frac{1}{7.10^4}\right).$$

Ainsi le pied carré vaut 1055 centimètres carrés, à moins d'un demi-centimètre carré près.

En divisant par 144 le résultat ci-dessus, on aurait la valeur du pouce carré,

$$1^{pq} = 7^{cmq},3 ;$$

et ainsi de suite.

3°. On propose d'évaluer la contenance du pied cube en litres.

De la valeur de la toise en mètres, $1^{m},9490$, on conclut

celle du pied en décimètres,

$$1^{p} = 3^{dm},248,$$

à moins d'une unité décimale du dernier ordre. La valeur du pied cube en litres sera donc exprimée approximativement par le cube de 3,248.

Ici nous avons $a < 4$, $\alpha < \frac{1}{10^3}$: l'erreur sur le cube sera moindre que $\frac{3.4^2}{10^3} = 0,048$, et, à fortiori, moindre que $\frac{1}{10}$. On ne pourra donc compter que sur les décilitres. En opérant par logarithmes, on trouve $(3,248)^3 = 34,264\ldots$ Donc,

$$1^{pc} = 34^{lit},3, \quad \text{à moins d'un décilitre.}$$

36. Quel sera le nombre des chiffres à employer pour a, afin que la puissance $n^{ième}$ soit calculée avec m chiffres exacts ?

Il suffit que l'erreur relative de a^n soit inférieure à $\frac{1}{10^m}$, et, par suite, que celle de a soit inférieure à $\frac{1}{n.10m}$. Donc, si le premier chiffre de a égale ou surpasse n, on calculera ce nombre avec $m+1$ chiffres; sinon il en faudra $m+2$. [On suppose le degré $n < 10$.]

Extraction des racines.

37. Nous avons donné (20) une *limite inférieure de l'erreur absolue d'un produit* de plusieurs facteurs approchés par défaut. Si l'on suppose ces facteurs égaux entre eux, on en conclura que l'erreur absolue commise sur a^n, lorsqu'au lieu de a on prend $a-\alpha$, est plus grande que $n\ (a-\alpha)^{n-1}\ \alpha$.

Cette limite va nous servir pour établir assez simplement une limite supérieure de l'erreur commise sur $\sqrt[n]{a}$, quand, au lieu de a, on emploie la valeur approchée $a-\alpha$.

En effet, on peut regarder a comme la puissance $n^{ième}$

de sa racine $n^{ième}$; donc, en désignant par x l'erreur commise sur $\sqrt[n]{a}$, on aura

$$\alpha > n \sqrt[n]{(a-\alpha)^{n-1}}\, x,$$

d'où

$$x < \frac{\alpha}{n \sqrt[n]{(a-\alpha)^{n-1}}}.$$

Les seules applications de cette limite, qui méritent d'être mentionnées, répondent à $n = 2$ et $n = 3$.

Pour $n = 2$, on a $$x < \frac{\alpha}{2\sqrt{a-\alpha}};$$

pour $n = 3$, on a $$x < \frac{\alpha}{3\sqrt[3]{(a-\alpha)^2}}.$$

Telles sont les limites supérieures de l'erreur absolue commise sur une racine carrée et cubique.

Passons à l'erreur relative; on déduit de la formule précédente :

$$\frac{x}{\sqrt[n]{a-\alpha}} < \frac{\alpha}{n(a-\alpha)}.$$

Or l'erreur relative $\frac{x}{\sqrt[n]{a}}$ est plus petite que $\frac{x}{\sqrt[n]{a-\alpha}}$; donc, à fortiori,

$$\frac{x}{\sqrt[n]{a}} < \frac{\alpha}{n(a-\alpha)}.$$

L'erreur relative de la racine n^{ième} *d'un nombre est moindre que la* n^{ième} *partie de l'erreur relative de ce nombre pris avec sa valeur approchée par défaut.*

Dans le calcul de cette limite, on pourra remplacer $a - \alpha$ par un nombre plus petit, par exemple le nombre d'unités du premier chiffre de a.

38. Quel sera le nombre des chiffres à employer pour a,

afin que sa racine $n^{ième}$ soit calculée avec m chiffres exacts?

Il suffit que l'erreur relative de la racine soit moindre que $\frac{1}{10^m}$, et, par suite, on doit avoir

$$\frac{\alpha}{a-\alpha} < \frac{n}{10^m}, \quad \text{d'où} \quad \frac{\alpha}{a} < \frac{n}{10^m + n}.$$

Or, dans les applications, n sera toujours moindre que 10^m, et d'ailleurs n est au moins égal à 2; donc il suffira qu'on ait

$$\frac{\alpha}{a} < \frac{2}{10^m + 10^m} \quad \text{ou} \quad \frac{\alpha}{a} < \frac{1}{10^m},$$

condition remplie si a présente $m + 1$ chiffres exacts.

Ainsi, *pour que* $\sqrt[n]{a}$ *soit calculée avec* m *chiffres, il suffira que* a *soit calculé par défaut avec* m + 1 *chiffres.*

En faisant usage des premiers chiffres de a et de $\sqrt[n]{a}$, lesquels sont souvent connus immédiatement, on peut abaisser la limite précédente. En effet, si k désigne le premier chiffre à gauche de $\sqrt[n]{a}$, il suffit que l'on ait

$$\frac{\alpha}{a-\alpha} < \frac{n}{(k+1)\,10^{m-1}}, \quad \text{d'où} \quad \frac{\alpha}{a} < \frac{1}{\frac{k+1}{n}\,10^{m-1} + 1}.$$

Donc, si le premier chiffre de a surpasse $\frac{k+1}{n}$, ou même s'il est égal à ce nombre et que les chiffres suivants ne soient pas tous des zéros, on pourra se borner à calculer a avec m chiffres.

Quand il s'agit de calculer la racine carrée ou la racine cubique d'un nombre a avec m décimales, les règles ordinaires de l'arithmétique prescrivent de chercher la valeur de ce nombre avec $2m$ décimales dans le premier cas, avec $3m$ dans le second. On voit que ces règles entraînent le plus souvent une complication inutile.

39. *Exemples.*—1°. Calculer $\sqrt{\frac{3}{7}}$ avec trois décimales exactes.

Ici $m = 3$; on cherchera donc les *quatre* premiers chiffres décimaux de la fraction $\frac{3}{7}$, ou 0,4285, que l'on fera suivre de cinq zéros ; puis on appliquera la règle ordinaire. La racine cherchée est 0,753.

2°. Calculer $\sqrt{215 + \frac{4}{7}}$ avec deux décimales exactes.

Comme cette racine aura évidemment deux chiffres à sa partie entière, c'est *quatre* chiffres exacts qu'il s'agit de connaître en définitive. Ainsi $m = 4$; d'ailleurs le premier chiffre k de la racine est 1, d'où $\frac{k+1}{n} = 1$. Donc (38) il suffira de calculer le nombre $\left(215 + \frac{4}{7}\right)$ avec m ou *quatre* chiffres. Or on connaît déjà les trois premiers ; on cherchera seulement la première décimale de la fraction $\frac{4}{7}$, et, la faisant suivre de trois zéros, on extraira la racine carrée de 2155000 à une unité près. Cette racine est 1467. Enfin, on aura

$$\sqrt{215 + \frac{4}{7}} = 14,68, \quad \text{à moins de } \frac{1}{100}.$$

La règle ordinaire de l'arithmétique aurait prescrit de chercher les quatre premières décimales de $\frac{4}{7}$. Nous retrouverons encore cette simplification au n° 47 par d'autres considérations, et nous ferons ressortir ce qu'elle offre de général.

3°. Le nombre 23456 étant supposé fautif de moins d'une unité, combien de chiffres exacts comportera sa racine carrée ?

Il sera plus court de calculer ici l'erreur *absolue* de la racine. Cette erreur est moindre que $\frac{1}{2\sqrt{23455}}$, et, à fortiori, moindre que $\frac{1}{200}$; ainsi on pourra compter sur les deux premières décimales de la racine, aussi bien que si le nombre proposé eût été calculé avec quatre chiffres décimaux exacts. On trouvera

$$\sqrt{23456} = 153,16$$

par excès.

4°. Calculer avec trois chiffres exacts le côté du carré équivalent au cercle dont le rayon est 1 mètre.

Ce côté est exprimé par $\sqrt[2]{\pi}$; on a $m = 3$; on prendra π avec *trois* chiffres seulement, attendu que le premier chiffre k de la racine sera 1. La question est ramenée à extraire la racine carrée du nombre 3,14 avec deux décimales. On trouve

$$\sqrt{3,14} = 1,77.$$

Le côté du carré cherché est donc $1^{m},78$, à moins d'un centimètre.

5°. Calculer avec trois chiffres exacts l'expression $\sqrt{10+2\sqrt{5}}$.

(Ce radical composé se présente dans la formule du côté du pentédécagone régulier, qui est $\left[\frac{\sqrt{10+2\sqrt{5}}-\sqrt{3}(\sqrt{5}-1)}{4}\right]$, en prenant pour unité le rayon du cercle circonscrit.)

On mettra d'abord le radical proposé sous la forme $\sqrt{10+\sqrt{20}}$; et puisqu'on demande trois chiffres exacts dans le résultat final, on devra calculer $\sqrt{20}$ avec quatre chiffres. Soit donc $\sqrt{20} = 4,472$. La racine carrée de 14,472 est 3,80...; par conséquent 3,81 est la valeur cherchée.

6°. Sachant que l'aire d'un cercle est égale à 645 déci-

mètres carrés, à une unité près, on propose de calculer la valeur approchée du rayon. (Question posée au nº 2.)

L'expression du rayon est $\sqrt{\frac{645}{\pi}}$; or, *quelle que soit l'approximation du diviseur* π, le quotient $\frac{645}{\pi}$ est déjà affecté d'une erreur relative qui a pour limite $\frac{1}{600}$, à cause que le dividende est approché à une unité près (26). On ne saurait donc se proposer de calculer ce quotient avec *plus de trois chiffres exacts;* et encore faudra-t-il, pour cela, que l'on emploie une valeur suffisamment approchée de π. En prenant ce nombre avec trois décimales, par excès (3,142), l'erreur relative du quotient $\frac{645}{3,142}$ s'augmentera d'une quantité moindre que $\frac{1}{3000}$; somme toute, elle sera moindre que $\left(\frac{1}{600}+\frac{1}{3000}\right)$ ou $\frac{1}{500}$; et, par suite, comme le premier chiffre du quotient est 2, son erreur absolue sera moindre que $\frac{300}{500}$ ou 0,6. On trouve

$$\frac{645}{3,142}=205,2\ldots.$$

En négligeant le chiffre des dixièmes et les suivants, l'erreur qui affectera la partie entière 205 sera moindre que 0.3, et cette erreur, jointe à la précédente, donnera une somme moindre que 0,9. Donc enfin 205 est la valeur par défaut de $\frac{645}{\pi}$ à moins d'une unité. Il reste à extraire la racine carrée de 205, laquelle aura trois chiffres exacts (37).

$$\sqrt{205}=14,3\ldots.$$

Le rayon cherché sera donc $14^{\text{déc.}},4$; et il ne serait pas possible, sans changer les données, de l'obtenir avec une approximation décimale plus grande.

7°. On propose de calculer la valeur approchée, à une unité près, de $\sqrt{\frac{375300}{\pi}}$, le nombre 375300 étant considéré comme exact.

La partie entière de la racine aura évidemment trois chiffres. Par conséquent, on devra calculer le quotient $\frac{375300}{\pi}$ avec *quatre* chiffres exacts, et, de plus, il faudra que l'on sache si, réduit à ces quatre chiffres, ce quotient reste approché par défaut. A cet effet, on opérera comme si l'on devait connaître *cinq* chiffres exacts. Or on voit tout de suite que le premier chiffre k du quotient sera 1, en sorte que le premier chiffre du diviseur π surpasse $k+1$; donc (29) il suffira de *cinq* chiffres pour π : soit donc $\pi = 3,1416$. Les cinq premiers chiffres du quotient $\frac{375300}{3,1416}$ forment le nombre 11946; le cinquième étant plus petit que 9, on est sûr qu'en le négligeant on aura pour valeur approchée de ce quotient, 119400, à une centaine près : car

$$\frac{375300}{\pi} = 119400 + \varepsilon, \quad \varepsilon < 700 + 100 \quad \text{ou} \quad < 800.$$

(On aperçoit la nécessité de calculer, comme ci-dessus, un chiffre de plus qu'on n'en veut définitivement conserver; autrement, comme on doit faire deux erreurs successives et de même sens, on ne saurait pas si le quotient approché par défaut est 1194 ou 1195 centaines, à une centaine près. On avait déjà vu des exemples de cette précaution aux n°s 11 et 22.)

Il reste à extraire la racine carrée de 119400 à une unité près; cette racine est 345. On prendra définitivement 346 pour la valeur de $\sqrt{\frac{375300}{\pi}}$ approchée à moins d'une unité.

Les derniers exemples que nous venons de traiter montrent la marche à suivre dans les opérations complexes où

le résultat cherché exige plusieurs calculs approximatifs superposés, et lorsqu'on ne peut pas recourir aux logarithmes. On applique successivement à chaque opération la formule d'approximation qui lui est propre. L'ordre à suivre dans l'application de ces diverses limites est inverse, selon qu'il s'agit de répondre à la première ou à la deuxième question du n° 1. Dans le premier cas, on applique d'abord la limite qui convient à celle des opérations qui doit être effectuée la première. Dans le second cas, on commence par celle qui doit être effectuée la dernière. Mais quand il arrive, comme dans la question précédente, qu'il n'entre dans le calcul qu'une seule quantité dont la valeur soit approchée, toutes les autres étant prises avec leurs valeurs exactes, il est plus simple d'employer *la formule générale de l'approximation*, qui sera exposée plus loin (48).

Extraction de la racine carrée abrégée.

40. Supposons que l'on connaisse une première valeur approchée par défaut a de la racine carrée du nombre entier N (a sera, par exemple, la racine du plus grand carré contenu dans la dernière tranche à gauche de N, suivie d'autant de zéros qu'il y a de tranches moins une). En divisant N par a, on aura un quotient plus grand que a, qu'on peut représenter par $a+b$. Nous allons prouver que la racine carrée de N sera comprise entre a et $a+\frac{b}{2}$.

En effet, soit $a+x$ la valeur exacte de cette racine ; on aura en même temps

$$N=(a+x)^2=a^2+2ax+x^2,$$
$$N=a(a+b)=a^2+ab;$$

égalant ces deux valeurs de N, on en tire

$$(1) \qquad x=\frac{b}{2}-\frac{x^2}{2a}.$$

Cette équation montre que x est moindre que $\frac{b}{2}$; par

suite, $a+\frac{b}{2}$ est une valeur approchée par excès de la racine. De plus, on voit qu'en prenant $\frac{b}{2}$ à la place de x, on néglige $\frac{x^2}{2a}$; or x est moindre que $\frac{b}{2}$: donc l'erreur est moindre que $\frac{b^2}{8a}$. En résumé,

$\mathrm{a}+\frac{\mathrm{b}}{2}$ *est une valeur approchée par excès de* $\sqrt{\mathrm{N}}$, *à moins de* $\frac{\mathrm{b}^2}{8\mathrm{a}}$.

Nous avons supposé que a était approchée par défaut. Si le contraire avait lieu, on pourrait toujours poser $\frac{\mathrm{N}}{a}=a+b$; b serait négatif. Il serait encore vrai de dire que $\left(a+\frac{b}{2}\right)$ est approchée par excès; car on a, quel que soit le signe de b,

$$\left(a+\frac{b}{2}\right)^2=a^2+ab+\frac{b^2}{4}=\mathrm{N}+\frac{b^2}{4}>\mathrm{N}.$$

Mais *la limite supérieure de l'erreur,* $\frac{\mathrm{b}^2}{8\mathrm{a}}$, *n'aurait plus lieu;* car, x et b étant négatifs, l'équation (1) montre que *la valeur numérique de* x *est plus grande que celle de* $\frac{\mathrm{b}}{2}$; par conséquent, l'erreur commise serait plus grande que $\frac{b^2}{8a}$.

Pour avoir une limite de l'erreur qui convienne aux deux cas, supposons que la valeur a d'où l'on part soit approchée à moins de ε (ε sera moindre que $\frac{b}{2}$, si a est par défaut). La valeur numérique de x sera moindre que ε, et celle de $\frac{x^2}{2a}$ moindre que $\frac{\varepsilon^2}{2a}$; donc, en désignant par ε' l'erreur

que comporte $a+\frac{b}{2}$, on aura toujours

$$\varepsilon' < \frac{\varepsilon^2}{2a}, \quad \text{ou} \quad \varepsilon' < \varepsilon . \frac{\varepsilon}{2a}.$$

Ainsi ε' sera moindre que ε, et l'approximation ira croissant, si ε est moindre que $2a$, ou, ce qui revient au même, si a surpasse le tiers de la racine cherchée. Cette condition sera toujours remplie dans la pratique.

Pour approcher davantage de la racine, posons

$$a+\frac{b}{2}=a_1,$$

et, raisonnant sur a_1 comme sur a, nous serons conduits à diviser N par a_1, puis à retrancher a_1 du quotient. Soit $\frac{\text{N}}{a_1}-a_1=b_1$ (b_1 sera négatif); une nouvelle valeur approchée de la racine sera

$$a_1+\frac{b_1}{2}=a_2,$$

et, désignant par ε'' l'erreur qu'elle comporte, on aura

$$\varepsilon'' < \frac{\varepsilon'^2}{2a_1}, \quad \text{et, à fortiori,} \quad \varepsilon'' < \frac{\varepsilon'^2}{2a}.$$

En continuant ainsi, on aura des valeurs de plus en plus approchées.

41. L'approximation en décimales, qui est seule usitée, marche avec une régularité et une rapidité remarquables. En effet, supposons que a soit approchée à $\frac{1}{10}$ près, cette valeur étant d'ailleurs plus grande que 1; on aura

$$\varepsilon < \frac{1}{10}, \quad \varepsilon' < \frac{1}{10^2}; \quad \varepsilon'' < \frac{1}{10^4} \cdots$$

Le nombre de décimales exactes que présentent les valeurs

approchées

$$a,\quad a_1,\quad a_2,\ldots,$$

ira donc constamment en doublant. D'après cela, comme dans le calcul de $a_1 = a + \frac{b}{2}$, on peut seulement répondre de l'exactitude des deux premiers chiffres décimaux, on ne prendra le quotient $\frac{N}{a}$ qu'avec deux décimales, et, négligeant les suivantes, on en conclura $\frac{b}{2}$ à $\frac{1}{100}$ près. On fera ici une nouvelle erreur en sens contraire de l'erreur théorique ε', puisque $a + \frac{b}{2}$ est en excès; et, par suite, la valeur de a_1 réduite à deux décimales ne cessera pas d'être approchée de la racine à $\frac{1}{100}$ près. Seulement, *elle pourra se trouver par défaut.*

Le sens de l'erreur est mis en évidence, dans la seconde approximation, par le calcul du quotient $\frac{N}{a_1}$, où le diviseur a_1 est réduit à deux décimales. Si ce quotient surpasse a_1, on en conclut que a_1 est une valeur par défaut, et dans ce cas b_1 n'est plus négatif.

Pour continuer la méthode, on calculera le quotient $\frac{N}{a_1}$ avec quatre décimales seulement. En général, on prendra chaque quotient avec un nombre de chiffres décimaux double de celui du diviseur.

Appliquons cette marche au calcul approché de $\sqrt{2}$. En partant de la valeur 1,4, voici le tableau des opérations :

$$a = 1{,}4;\quad \frac{2}{1{,}4} = 1{,}42\ldots;\quad \frac{b}{2} = 0{,}01;$$

$$a_1 = 1{,}41;\quad \frac{2}{1{,}41} = 1{,}4184\ldots;\quad \frac{b_1}{2} = 0{,}0042;$$

$$a_2 = 1{,}4142;\quad \frac{2}{1{,}4142} = 1{,}41422712;\quad \frac{b_2}{2} = 0{,}00001356;$$

$$a_3 = 1{,}41421356.$$

Cette dernière valeur est approchée avec huit décimales exactes.

Ce qui précède suppose simplement $a > 1$; mais si l'on a

$$2a \geqq 10^n,$$

l'approximation marchera encore plus rapidement; car, en supposant toujours que a soit approchée à $\frac{1}{10}$ près de la racine, il vient

$$\varepsilon < \frac{1}{10}, \quad \varepsilon' < \frac{1}{10^{n+2}}, \quad \varepsilon'' < \frac{1}{10^{3n+4}} \cdots;$$

en sorte que le nombre des décimales exactes que chaque approximation fournit, surpasse au moins de n unités le double du nombre des décimales fournies par l'approximation précédente. Dans ce cas, si l'on partait d'une valeur a approchée à une unité près, la division de N par a fournirait tout de suite une valeur a_1 approchée à $\frac{1}{10^n}$ près; car on aurait

$$\varepsilon' < \frac{1}{2a} < \frac{1}{10^n}.$$

42. Si l'on demande *quelle doit être la valeur de* a *pour que, dès la première division,* $a + \frac{b}{2}$ *soit approchée à moins d'une unité*, on posera la condition

$$\frac{x^2}{2a} < 1;$$

or, en désignant par p le nombre des chiffres entiers qui restent à trouver, on a

$$x < 10^p, \quad x^2 < 10^{2p}.$$

La condition précédente revient donc à

$$2a \geqq 10^{2p},$$

et, comme la valeur approchée a se termine par p zéros, *il suffira que l'on connaisse déjà* p + 1, *ou plus de la moitié des chiffres de la racine,* mais cela ne sera pas toujours nécessaire.

Remarque. — Nous retrouvons ici la condition connue, pour qu'en partant d'une valeur approchée a, on puisse obtenir tous les autres chiffres entiers de la racine par une simple division. Et, en effet, la règle qu'on donne ordinairement et qui consiste à calculer la partie entière du quotient de $(N - a^2)$ par $2a$, ne diffère pas, au fond, de celle que nous venons d'exposer, puisque l'on a

$$\frac{N - a^2}{2a} = \frac{1}{2}\left(\frac{N}{a} - a\right) = \frac{b}{2}.$$

Mais elle se prête moins bien au calcul de la racine par approximations successives; le procédé que nous suivons évite la formation du carré de chaque valeur approchée telle que a, et celle du carré de chaque quotient, opérations assez longues et nécessaires quand on veut appliquer la règle ordinaire.

Remarquons de plus que chaque division sert de preuve et de correctif à la précédente, et manifeste le sens de l'approximation déjà obtenue. Car la vraie racine est nécessairement comprise entre le diviseur et le quotient correspondant.

43. *Exemples.* — 1°. On propose de calculer approximativement la racine carrée du nombre 396854.

La racine du plus grand carré contenu dans 39 étant 6, la racine cherchée est comprise entre 600 et 700. Partons

de $a = 600$; il vient

$$\frac{396854}{600} = 661,\ldots,$$

d'où

$$b = 61,\ldots,\quad \frac{b}{2} = 30,\ldots,\quad a + \frac{b}{2} = 630,\ldots,$$

$$\varepsilon' < \frac{b^2}{8a} < \frac{(31)^2}{2.600} = \frac{961}{1200} < 1.$$

Donc 630 est déjà la racine à une unité près. Dans l'approximation suivante, on ira jusqu'aux millièmes; car on a $\varepsilon'' < \frac{1}{2a} < \frac{1}{10^3}$. On trouve

$$\frac{396854}{630} = 629,926\ldots.$$

Ce quotient étant moindre que le diviseur, on en conclut que 630 est une valeur par excès; de plus on a

$$b_1 = -0,074,\quad \frac{b_1}{2} = -0,037,$$

$$a_1 + \frac{b_1}{2} = 630 - 0,037 = 629,963;$$

ainsi la racine, à un millième près, est 629,963 (*).

Dans l'approximation suivante, on aura

$$\varepsilon''' < \frac{\varepsilon''^2}{2a} < \frac{1}{10^9};$$

(*) On évite les deux soustractions qu'exigent le calcul de b_1 et celui de la somme $a_1 + \frac{b_1}{2}$, en remarquant que

$$a_1 + \frac{b_1}{2} = 630 + \frac{629,926 - 630}{2} = 629 + \frac{631,926 - 630}{2} = 629 + \frac{1,926}{2}.$$

Il suffit donc d'*ajouter au diviseur, diminué d'une unité de l'ordre de son dernier chiffre, la moitié de la différence entre le diviseur et le quotient, différence dont le premier chiffre à gauche est augmenté de* 10. Ainsi, après avoir écrit 629, on dira : La moitié de 19 est 9, etc.

on poussera donc le quotient $\frac{396854}{629,963}$ jusqu'à la neuvième décimale, d'où l'on conclura une valeur approchée de la racine à moins d'une unité du neuvième ordre décimal. On voit combien l'approximation est rapide. On abrégerait encore ces calculs en combinant la méthode avec le procédé de la division abrégée exposé au n° 31.

2°. Proposons-nous encore le calcul approché de $\sqrt{5}$. Voici le tableau des opérations en partant de $a = 2$.

$a = 2 \quad \frac{5}{2} = 2,5, \quad \frac{b}{2} = 0,25, \quad \varepsilon' < \frac{0,25}{16} < \frac{1}{10};$

$a + \frac{b}{2} = 2,2$, valeur approchée à $\frac{1}{10}$ près;

$a_1 = 2,2, \quad \frac{5}{2,2} = 2,27\ldots, \quad \frac{b_1}{2} = 0,03\ldots, \quad \varepsilon'' < \frac{1}{10^2};$

$a_1 + \frac{b_1}{2} = 2,23$, valeur approchée à $\frac{1}{100}$ près (a_1 est en défaut);

$a_2 = 2,23, \quad \frac{5}{2,23} = 2,2421\ldots, \quad \frac{b_2}{2} = 0,0060\ldots, \quad \varepsilon''' < \frac{1}{10^4};$

$a_2 + \frac{b_2}{2} = 2,2360$, valeur approchée à $\frac{1}{10000}$ près (a_2 est en défaut);

$a_3 = 2,2360, \quad \frac{5}{2,2360} = 2,23613595\ldots, \quad \varepsilon^{\text{IV}} < \frac{1}{10^8};$

$a_4 = 2,23606797$, valeur approchée avec huit décimales exactes.

Si l'on s'arrête à cette approximation, et qu'on veuille la vérifier et en reconnaître le sens, on calculera le quotient suivant, $\frac{5}{2,23606797}$, en se bornant aux huit premières décimales. On trouvera 2,23606798... ; les sept premières décimales sont les mêmes que celles du diviseur, et la huitième est plus grande d'une unité. Par conséquent, la valeur a_4 est bien approchée à moins de $\frac{1}{10^8}$, et elle l'est par défaut.

44. C'est ici le lieu d'exposer quelques formules d'approximation auxquelles on a souvent recours dans l'extraction des racines carrées et cubiques.

Si l'on fait le carré de $1+\frac{\varepsilon}{2}$, on trouve $1+\varepsilon+\frac{\varepsilon^2}{2}$, résultat qui diffère peu de $1+\varepsilon$, quand ε est une petite fraction de l'unité. On est ainsi conduit à prendre $1+\frac{\varepsilon}{2}$ pour valeur approchée de $\sqrt{1+\varepsilon}$.

L'erreur commise est

$$\alpha = 1+\frac{\varepsilon}{2}-\sqrt{1+\varepsilon},$$

ou bien, en multipliant et divisant par la somme

$$1+\frac{\varepsilon}{2}+\sqrt{1+\varepsilon},$$

$$\alpha = \frac{\frac{\varepsilon^2}{4}}{1+\frac{\varepsilon}{2}\sqrt{1+\varepsilon}}, \quad \text{d'où} \quad \alpha < \frac{\varepsilon^2}{8}.$$

Concluons de là que, *pour extraire la racine carrée d'un nombre* $1+\varepsilon$, *qui surpasse peu l'unité, il suffit d'ajouter à l'unité la moitié de l'excès* ε; *et que le résultat est approché par excès, à moins de* $\frac{\varepsilon^2}{8}$.

Par exemple, on aura

$$\sqrt{1,000694} = 1+\frac{0,000694}{2} = 1,000347,$$

à moins de $\frac{7^2}{8.10^8}$ ou de $\frac{1}{10^7}$. L'erreur commise n'influera donc pas sur le dernier chiffre décimal de la valeur approchée 1,000347. En général, s'il y a p zéros entre l'unité et le premier chiffre décimal significatif du nombre dont on demande la racine carrée, l'erreur commise en prenant

$1 + \frac{\varepsilon}{2}$ pour $\sqrt{1 + \varepsilon}$ n'influera pas sur les $2p$ premiers chiffres décimaux de la racine.

45. Si l'on fait le cube de $1 + \frac{\varepsilon}{3}$, on trouve $1 + \varepsilon + \frac{\varepsilon^2}{3} + \frac{\varepsilon^3}{27}$, résultat qui ne diffère de $1 + \varepsilon$ que par des termes d'ordre supérieur à ε. On est ainsi conduit à prendre $1 + \frac{\varepsilon}{3}$ pour valeur approchée de $\sqrt[3]{1 + \varepsilon}$.

L'erreur commise est

$$\alpha = 1 + \frac{\varepsilon}{3} - \sqrt[3]{1 + \varepsilon},$$

ou bien, en multipliant et divisant par une quantité telle, que le numérateur devienne la différence des cubes de $1 + \frac{\varepsilon}{3}$ et de $\sqrt[3]{1 + \varepsilon}$ (*),

$$\alpha = \frac{\left(1 + \frac{\varepsilon}{3}\right)^3 - (1 + \varepsilon)}{\left(1 + \frac{\varepsilon}{3}\right)^2 + \left(1 + \frac{\varepsilon}{3}\right)\sqrt[3]{1 + \varepsilon} + (\sqrt[3]{1 + \varepsilon})^2},$$

$$= \frac{\frac{\varepsilon^2}{3} + \frac{\varepsilon^3}{27}}{\left(1 + \frac{\varepsilon}{3}\right)^2 + \left(1 + \frac{\varepsilon}{3}\right)\sqrt[3]{1 + \varepsilon} + (\sqrt[3]{1 + \varepsilon})^2};$$

d'où

$$\alpha < \frac{\varepsilon^2}{3}\left(\frac{1}{3} + \frac{1}{27}\right) < \frac{\varepsilon^2}{8}.$$

(*) Si l'on pose en général $\left(1 + \frac{\varepsilon}{n}\right) = a$ et $\sqrt[n]{1 + \varepsilon} = b$, la transformation dont il s'agit est un cas particulier de la formule connue,

$$a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \ldots + b^{n-1});$$

d'où

$$a - b = \frac{a^n - b^n}{a^{n-1} + a^{n-2}b + a^{n-3}b^2 + \ldots + b^{n-1}}.$$

Ainsi, *pour extraire la racine cubique d'un nombre* $1+\varepsilon$ *qui surpasse peu l'unité, il suffit d'ajouter à l'unité le tiers de* ε. *Le résultat est approché par excès à moins de* $\frac{\varepsilon^2}{8}$.

Ces formules s'étendraient aux racines de tous les degrés. Mais il ne faut pas chercher, par des transformations analogues aux précédentes, une limite supérieure de l'erreur que l'on fait en remplaçant $\sqrt[n]{1+\varepsilon}$ par $1+\frac{\varepsilon}{n}$, attendu que la formule générale des approximations, qui sera exposée plus loin (74), y conduira d'une manière beaucoup plus simple.

46. *Lorsque deux nombres entiers* a, b *ont un même nombre de chiffres, et que plus de la moitié de ces chiffres sont communs à partir de la gauche, on a*

$$\sqrt{a}-\sqrt{b}<\frac{1}{2}\quad \text{et}\quad \frac{a+b}{2}-\sqrt{ab}<\frac{1}{8};$$

c'est-à-dire qu'*on peut remplacer la racine carrée de* a *par celle de* b, *sans que l'erreur atteigne une demi-unité; et la moyenne arithmétique entre* a *et* b *par leur moyenne géométrique, sans que l'erreur s'élève à un huitième d'unité.*

En effet, le nombre des chiffres de la différence $a-b$ étant, par hypothèse, moindre que la moitié du nombre des chiffres de b, et, d'autre part, le nombre des chiffres d'un carré étant au plus égal au double du nombre des chiffres de sa racine, on aura

$$(a-b)^2<b,\quad \text{d'où}\quad a-b<\sqrt{b},$$

ou bien encore

$$(\sqrt{a}-\sqrt{b})(\sqrt{a}+\sqrt{b})<\sqrt{b}.$$

Il en résulte

$$\sqrt{a}-\sqrt{b}<\frac{\sqrt{b}}{\sqrt{a}+\sqrt{b}},$$

et, à fortiori,

$$\sqrt{a}-\sqrt{b}<\frac{1}{2}. \qquad \text{C. Q. F. D.}$$

Enfin, élevant au carré les deux membres, il vient

$$a+b-2\sqrt{ab}<\frac{1}{4};$$

d'où

$$\frac{a+b}{2}-\sqrt{ab}<\frac{1}{8}. \qquad \text{C. Q. F. D.}$$

D'après l'inégalité $\sqrt{a}-\sqrt{b}<\frac{1}{2}$, on voit qu'on pourra, sans altérer d'une demi-unité la racine carrée d'un nombre, modifier à volonté des chiffres sur la droite, les réduire à des zéros par exemple, pourvu que les modifications ne portent pas sur la première moitié des chiffres à partir de la gauche.

Remarque. — Nous avons supposé que a et b étaient des nombres entiers. Mais la formule précédente s'applique aussi à deux nombres décimaux, dont les parties décimales ont le même nombre de chiffres. Seulement, l'unité simple sera remplacée par l'unité décimale du dernier ordre : ceci demande quelque éclaircissement.

Lorsque a et b sont des nombres décimaux, *il ne faut pas croire que la différence $\sqrt{a}-\sqrt{b}$ soit moindre qu'une demi-unité du dernier ordre décimal.* En effet, soit p le nombre des chiffres décimaux de a ou de b, et posons

$$a.10^p=a', \quad b.10^p=b'.$$

a' et b' étant des nombres entiers qui ont plus de la moi-

tié de leurs chiffres communs, on aura, par ce qui précède,

$$\sqrt{a'} - \sqrt{b'} < \frac{1}{2},$$

ou

$$\sqrt{a.10^p} - \sqrt{b.10^p} < \frac{1}{2};$$

et, divisant par $\sqrt{10^p}$,

$$\sqrt{a} - \sqrt{b} < \frac{1}{2\sqrt{10^p}}.$$

Ainsi, pour avoir la limite de l'erreur commise en prenant $\sqrt{a}$ pour $\sqrt{b}$, il faut diviser $\frac{1}{2}$ par la racine carrée de la puissance de 10 marquée par le nombre des chiffres décimaux.

Si, par exemple, a et b ont chacun quatre chiffres décimaux, et plus de la moitié de leurs chiffres communs sur la gauche, l'erreur commise en confondant $\sqrt{a}$ avec $\sqrt{b}$ sera moindre qu'un demi-centième, et non pas moindre qu'une demi-unité de l'ordre des dix-millièmes.

Au contraire, *la limite supérieure de l'erreur commise en confondant la moyenne arithmétique avec la moyenne géométrique est toujours un huitième de l'unité du dernier ordre décimal.* Cela tient à ce que l'élévation au carré des deux membres de l'inégalité précédente fait disparaître $\sqrt{10^p}$, et qu'il vient

$$\frac{a+b}{2} - \sqrt{ab} < \frac{1}{8.10^p}.$$

47. *Exemples.* — Soit proposé d'extraire la racine carrée du nombre $\left(215 + \frac{4}{7}\right)$ avec deux décimales, question déjà traitée au n° 39.

Pour plus de clarté, reprenons d'abord le raisonnement ordinaire de l'arithmétique.

Soit $\frac{x}{100}$ le plus grand nombre de centièmes dont le carré est inférieur à $\left(215 + \frac{4}{7}\right)$; on doit avoir

$$\frac{x^2}{100^2} < 215 + \frac{4}{7} < \frac{(x+1)^2}{100^2};$$

d'où

$$x^2 < 2150000 + \frac{40000}{7} < (x+1)^2.$$

Extrayant l'entier contenu dans $\frac{40000}{7}$, et l'ajoutant à 2150000, il vient

$$x^2 < 2155714 + \frac{2}{7} < (x+1)^2.$$

Maintenant, on fait voir que la fraction $\frac{2}{7}$ peut être supprimée sans que la double inégalité soit troublée. En effet, le nombre $2155714 + \frac{2}{7}$ étant moindre que $(x+1)^2$, on a, à fortiori, $2155714 < (x+1)^2$; et, d'autre part, x^2 qui doit être entier et moindre que $2155714 + \frac{2}{7}$, est au plus égal à 2155714. On peut donc écrire

$$x^2 \leqq 2155714 < (x+1)^2;$$

d'où l'on conclut enfin que x n'est autre que la racine du plus grand carré contenu dans 2155714. Par suite, on extrait la racine carrée de ce nombre à une unité près, et il reste à la diviser par 100 pour avoir la racine cherchée.

Voilà bien le calcul tel que le prescrit l'arithmétique ordinaire. Mais est-il nécessaire de connaître les sept chiffres du nombre entier 2155714?

Il résulte du théorème précédent que les quatre premiers

suffisent : car on a

$$\sqrt{2155714} - \sqrt{2155000} < \frac{1}{2}.$$

Donc, dans la pratique, on pourra se borner à chercher le premier chiffre 5 du quotient $\frac{40000}{7}$, ou, ce qui revient au même, à *chercher le chiffre des dixièmes de la fraction* $\frac{4}{7}$, *convertie en décimales*. On l'écrira à la droite du nombre entier donné 215; ce qui donnera 2155 qu'on fera suivre de trois zéros. Enfin on extraira la racine de 2155000 à moins d'une unité.

Cette racine est 1467 : mais on doit prendre 1468, attendu qu'il y a ici *deux* erreurs superposées : l'une, plus petite que $\frac{1}{2}$, provenant de ce que l'on a substitué 2155000 à 2155714; l'autre, plus petite que 1, provenant de ce que 1467 n'est pas la racine exacte de 2155000. Or il ne faut pas que ces deux erreurs soient de même sens; autrement elles s'ajouteraient et leur somme pourrait dépasser 1. C'est précisément ce qui arriverait dans le cas qui nous occupe, car on peut s'assurer que la racine du plus grand carré contenu dans 2155714 est 1468 et non 1467.

En général, étant donné un nombre entier accompagné d'une fraction ordinaire, dont on demande la racine carrée avec n décimales, quel sera le nombre de chiffres nécessaire à calculer dans la conversion de cette fraction ordinaire en décimales? Distinguons deux cas :

1°. Le nombre des chiffres de la partie entière du nombre donné est impair $2p+1$. Ce nombre multiplié par 10^{2n} présentera $2p+1+2n$ chiffres, dont les $p+n+1$ premiers seront seuls utiles à connaître. Or on a

$$p+n+1 = 2p+1+(n-p),$$

et déjà les $2p+1$ premiers chiffres sont donnés par la partie

entière du nombre proposé ; donc *il reste à calculer* $n-p$ *chiffres* seulement. C'est la conversion de la fraction ordinaire en décimales qui les fournira.

(Nous supposons $n > p$; autrement il n'y aurait pas lieu de tenir compte de la fraction ordinaire qui accompagne l'entier, ce dernier fournissant déjà autant ou plus de chiffres qu'il n'en est besoin.)

On demande, par exemple,

$$\sqrt{34567+\frac{8}{13}} \text{ avec trois décimales.}$$

Ici $n=3$, $p=2$; d'où $n-p=1$.

Il suffira de chercher le premier chiffre décimal 6 de $\frac{8}{13}$, puis d'extraire la racine carrée du nombre 3456760000, à une unité près.

Si l'on avait demandé la racine carrée de $\left(34567+\frac{8}{13}\right)$ avec *deux* décimales seulement, on aurait eu $n-p=0$, et par conséquent la fraction $\frac{8}{13}$ ne serait intervenue en rien dans le calcul.

Enfin, si la racine carrée n'avait dû être calculée qu'avec *une* décimale, on aurait eu $n-p=-1$, ce qui aurait signifié que la partie entière 34567 présentait déjà un chiffre de trop et qu'on pouvait se borner à prendre 3456 suivi de trois zéros.

2°. Le nombre des chiffres de la partie entière du nombre donné est pair, $2p$. Ce nombre multiplié par 10^{2n}, présentera $2p+2n$ chiffres, dont les $p+n+1$ premiers seront seuls utiles à connaître. Or on a

$$p+n+1=2p+(n-p+1);$$

et déjà les $2p$ premiers chiffres sont donnés par la partie entière du nombre; donc *il reste à calculer* $n-p+1$

chiffres, que l'on prendra sur la fraction ordinaire convertie en décimales.

On demande, par exemple,

$$\sqrt{2345 + \frac{6}{7}} \text{ avec deux décimales.}$$

$$n = 2, \quad p = 2; \quad \text{d'où} \quad n - p + 1 = 1.$$

On ne demandera donc qu'un chiffre à la conversion de $\frac{6}{7}$ en décimales, etc.

En résumé, *selon que le nombre des chiffres de la partie entière du nombre proposé sera égal à* $2p+1$ *ou à* $2p$, *on aura* $n-p$ *ou* $n-p+1$ *chiffres à chercher dans la conversion de la fraction ordinaire en décimales.*

Reprenons encore l'exemple du n° **44** : il s'agissait d'extraire la racine carrée du nombre 1,000694. Or on peut écrire

$$\sqrt{1,000694} = \sqrt{1,000694 \times 1,000000},$$

et, remplaçant la moyenne géométrique par la moyenne arithmétique, il viendra

$$\sqrt{1,000694} = \frac{1,000694 + 1}{2} = 1 + \frac{0,000694}{2}$$
$$= 1,000347,$$

résultat conforme à celui déjà obtenu.

Formule générale de l'approximation.

48. Les formules que nous venons d'exposer pour chacune des six opérations de l'Arithmétique peuvent toutes se déduire d'une formule générale qui comprend, en outre, les cas où il s'agit d'opérations transcendantes, telles que les calculs de logarithmes, de lignes trigonométriques, etc.

Rappelons d'abord quelques propositions relatives aux dérivées des fonctions continues d'une seule variable.

I. *Si la dérivée* f$'$(x) *de la fonction continue quelconque* f(x), *est positive pour la valeur* x = a, *la fonction est croissante,* c'est-à-dire que l'on a

$$f(a+h) > f(a),$$

h étant un nombre positif très-petit.

En effet, la dérivée $f'(a)$ étant la limite du rapport $\frac{f(a+h)-f(a)}{h}$ lorsque h tend vers zéro, on a

$$\frac{f(a+h)-f(a)}{h} = f'(a) + \varepsilon,$$

ε ayant pour limite zéro, en même temps que h. Pour de très-petites valeurs de h, le second membre prend le signe de $f'(a)$, c'est-à-dire le signe $+$; donc il en est de même du premier membre, et l'on a $f(a+h) > f(a)$.

C. Q. F. D.

II. On démontre de même que *si la dérivée* f$'$(x) est *négative pour* x = a, *la fonction est décroissante*, c'est-à-dire que l'on a $f(a+h) < f(a)$.

III. Il suit immédiatement de là que *si la dérivée* f$'$(x) *reste constamment positive* (ou *négative*), *tandis que* x *croît d'une manière continue depuis une valeur* α *jusqu'à* β, *la fonction* f(x) *sera continuellement croissante* (ou *décroissante*) *dans le même intervalle.*

Remarque. — Le raisonnement précédent exige que la dérivée ne s'annule pas pour $x = a$. Cependant, si la dérivée se réduisait à zéro pour une ou plusieurs valeurs de x comprises entre α et β, *sans cesser d'être constamment positive* (ou *négative*) *pour les autres valeurs de* x, les conclusions précédentes subsisteraient. En effet, supposons que l'on ait

$$f'(x) \geqq 0$$

lorsque x croît de α à β, et considérons la fonction

$$f(x) + kx,$$

k désignant une constante arbitraire, positive. La dérivée $f'(x) + k$ restera constamment positive, *sans passer par zéro*, pour toute valeur de x comprise entre α et β. Donc la fonction $f(x) + kx$ sera croissante, et l'on aura

$$f(a+h) + k(a+h) > f(a) + ka,$$

ou

$$f(a+h) - f(a) + kh > 0.$$

Il en résulte qu'on ne saurait avoir

$$f(a+h) - f(a) = -\delta^2,$$

quelque petite que soit la quantité δ; car on pourrait toujours assigner à l'indéterminée k une valeur assez petite pour que l'inégalité

$$-\delta^2 + kh > 0$$

ne subsistât plus. Ainsi, il est impossible que la fonction $f(x)$ aille en décroissant. On démontrerait de même, en donnant à k une valeur négative, que si la dérivée est constamment négative ou nulle, la fonction ne va jamais en croissant.

49. *Corollaire* I. — La dérivée $f'(x)$ ne saurait être nulle pour toute valeur de x, comprise entre α et β, sans que la fonction $f(x)$ se réduise à une constante, dans ce même intervalle. En effet, de ce que la dérivée n'est jamais négative, on conclura, comme précédemment, à l'aide du terme auxiliaire kx, que la fonction ne va pas en décroissant; et de ce que la dérivée n'est jamais positive on conclura que la fonction ne va pas en croissant; donc elle est constante pour toute valeur de x comprise entre α et β.

Corollaire II. — On dit qu'une valeur $x = a$ rend $f(x)$

maximum, quand on a la double inégalité

$$f(a-h) < f(a) > f(a+h),$$

h étant aussi petit qu'on voudra ; et si les deux signes d'inégalité sont renversés, en sorte que l'on ait

$$f(a-h) > f(a) < f(a+h),$$

on dit que $x=a$ rend $f(x)$ *minimum*. Cela posé, une condition nécessaire et suffisante pour que $x=a$ donne un maximum ou un minimum, c'est *que la dérivée* $f'(x)$ *change de signe, lorsque* x *croît de* $a-h$ *à* $a+h$. Si le changement de signe a lieu de $+$ à $-$, c'est que $f(x)$, d'abord croissante en deçà de a, devient décroissante au delà ; par conséquent il y a maximum. Si le changement de signe a lieu de $-$ à $+$, il y a au contraire minimum.

Quand la dérivée $f'(x)$ demeure finie et continue pour toutes les valeurs de x que l'on considère, elle ne peut changer de signe qu'en passant par zéro. On obtient donc alors toutes les valeurs de x susceptibles de donner des maxima ou des minima, en résolvant l'équation

$$f'(x)=0;$$

toutefois, quand on a trouvé une racine a de cette équation, il faut, avant de prononcer qu'elle répond à un maximum ou à un minimum, s'assurer que $f'(x)$ change bien de signe entre $a-h$ et $a+h$, h étant aussi petit qu'on voudra.

50. Ces préliminaires posés, nous pouvons établir la formule générale de l'approximation. Elle est comprise dans le théorème suivant :

Si la fonction f(x) *est telle, que sa dérivée reste finie et continue pour toute valeur de* x *comprise entre* a *et* a + h, *la différence* f(a + h) — f(a) *est égale au produit de* h *par la valeur que prend la dérivée* f'(x) *pour une certaine va-*

leur de x *intermédiaire entre* a *et* a + h ; *en sorte qu'on a*

$$(1) \qquad f(a+h)-f(a)=hf'(a+\theta h); \quad [0<\theta<1].$$

(h est ici un nombre donné qui peut n'être pas très-petit.)

Supposons d'abord h positif; soit M la plus grande des valeurs que prend $f'(x)$, quand x croît d'une manière continue depuis a jusqu'à $a+h$. Nous aurons, pour toutes ces valeurs de x,

$$M-f'(x) \geqq 0.$$

Or $M-f'(x)$ est évidemment la dérivée, par rapport à x, de la fonction

$$f(a+h)-f(x)-(a+h-x)M;$$

donc cette fonction est croissante ; et, comme elle s'annule pour la plus grande valeur de x, $a+h$, elle est constamment négative entre a et $a+h$. Si l'on y fait $x=a$, on aura donc

$$f(a+h)-f(a)-hM<0.$$

En désignant par m la plus petite des valeurs que prend la dérivée $f'(x)$ entre a et $a+h$, on démontrera de même l'inégalité

$$f(a+h)-f(a)-hm>0.$$

La différence $f(a+h)-f(a)$ est donc comprise entre hm et hM ; ou, ce qui revient au même, elle est égale au produit de h par une grandeur moyenne entre m et M.

Nous avons supposé h positif. Dans le cas contraire, $a+h$ étant moindre que a, on fera croître x depuis $a+h$ jusqu'à a, et les deux inégalités précédentes ne feront que changer de sens. La conclusion demeurera donc la même.

Ce qui précède n'exige pas que la dérivée $f'(x)$ soit *continue* entre a et $a+h$; il suffit qu'elle reste *finie*. Maintenant, si on la suppose continue, elle passera par tous les états de grandeur entre m et M. Donc il existera

une certaine valeur de x comprise entre a et $a+h$, que l'on peut désigner par $a+\theta h$, et pour laquelle $f'(x)$ deviendra précisément égale à la grandeur intermédiaire entre m et M dont il est question. L'équation (1) est donc démontrée.

51. *Remarque.* — L'équation (1) admet une interprétation géométrique bien simple.

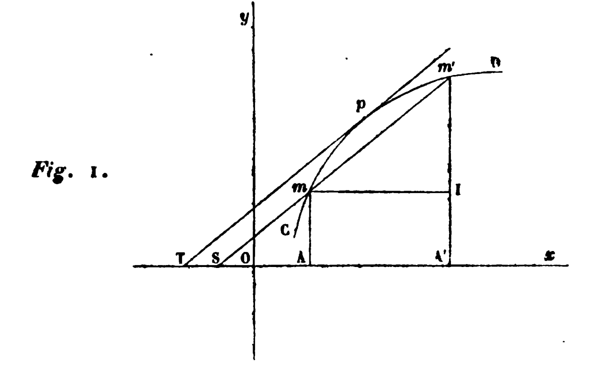
Posons

$$y=f(x),$$

et regardons y comme l'ordonnée variable d'une courbe dont x désignera l'abscisse.

Soit CmD cette courbe (*fig.* 1), rapportée aux deux axes

Fig. 1.



rectangulaires Ox, Oy; m et m' les deux points qui ont pour coordonnées

$$\begin{aligned} OA &= a, \qquad m\,A = f(a),\\ OA' &= a+h, \; m'A' = f(a+h). \end{aligned}$$

Si l'on mène mI parallèle à l'axe des x, le rapport

$$\frac{m'I}{mI} \quad \text{ou} \quad \frac{f(a+h)-f(a)}{h},$$

sera la tangente trigonométrique de l'angle $m'm$I que fait la sécante mm' avec l'axe des x. On sait d'ailleurs que la dérivée $f'(x)$ représente, pour chaque valeur de x, la tangente trigonométrique de l'angle que fait avec l'axe des x la tangente au point de la courbe CmD qui répond à cette abscisse.

Or, d'après l'équation (1), le rapport $\frac{f(a+h)-f(a)}{h}$ est égal à $f'(a+\theta h)$. Donc cette équation signifie qu'*il existe toujours sur la courbe, entre m et m', un point où la tangente est parallèle à la sécante mm'*. Il peut d'ailleurs y avoir *plusieurs* points, tels que p, où la tangente pT jouisse de cette propriété.

52. *Corollaire* I. — La dérivée $f'(x)$ ne saurait être nulle pour toute valeur de x comprise entre deux nombres donnés α et β, sans que la fonction $f(x)$ se réduise à une constante dans ce même intervalle. En effet, $f'(a+\theta h)$ étant alors nulle quel que soit h entre zéro et $\beta-\alpha$, il suit de l'équation (1) que l'on a constamment $f(\alpha+h)=f(\alpha)$, c'est-à-dire que la fonction ne varie pas avec x. Cette proposition avait déjà été prouvée (49).

Corollaire II. — L'équation (1) fournit l'expression de l'erreur que l'on commet dans la *règle de fausse position* ou *règle des parties proportionnelles*, quand *on suppose les accroissements d'une fonction proportionnels aux accroissements très-petits de la variable*. En effet, il résulte d'abord de la définition même d'une dérivée, que l'on a

$$f(a+h)-f(a)=hf'(a)+h\varepsilon,$$

ε ayant zéro pour limite en même temps que h. Le terme $h\varepsilon$ est donc infiniment petit par rapport à $hf'(a)$ [pourvu toutefois que $f'(a)$ ne soit pas nulle]; et par conséquent l'accroissement $f(a+h)-f(a)$ est sensiblement égal à $hf'(a)$ pour de très-petites valeurs de h, c'est-à-dire qu'il est proportionnel à h.

De plus, il résulte de l'équation (1) que ce même accroissement peut s'écrire :

$$f(a+h)-f(a)=hf'(a)+h[f'(a+\theta h)-f'(a)];$$

et par suite on voit que l'erreur commise en supposant la proportionnalité ci-dessus est une fraction de h marquée par $f'(a+\theta h)-f'(a)$. Nous l'apprécierons plus loin sur divers exemples.

53. L'hypothèse de la proportionnalité des accroissements d'une fonction aux accroissements très-petits de la variable, est utile dans un grand nombre de cas, où l'on doit étudier l'effet produit sur une grandeur, par une petite variation de l'un des éléments dont cette grandeur dépend.

Par exemple, quand on veut déterminer la hauteur AB (*fig.* 2) d'un édifice, par la trigonométrie, on mesure une

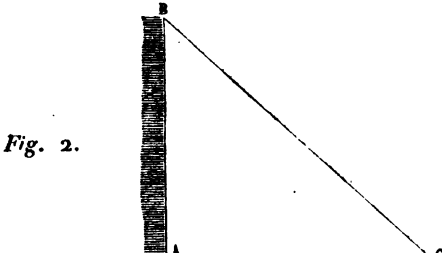


Fig. 2.

base AC à partir du pied de la verticale, puis l'angle BCA. Désignons par C cet angle, par b la base et par h la hauteur cherchée. Si l'angle C était connu exactement, on aurait

$$h=b\tan C.$$

Mais il est impossible d'éviter une certaine erreur absolue dans la mesure de l'angle C. Cette erreur est due, en partie,

à l'imperfection de la graduation de l'instrument qui ne permet pas de lire le nombre exact de degrés, minutes et secondes de l'angle; en partie, à la difficulté de pointer juste vers le point B. Désignons cette erreur par α, en sorte que l'angle observé sera $C+\alpha$. L'erreur absolue qui en résultera sur la hauteur cherchée, sera

$$\varepsilon = b[\tang(C+\alpha) - \tang C],$$

et l'erreur relative correspondante,

$$\frac{\varepsilon}{h} = \frac{\tang(C+\alpha) - \tang C}{\tang C}.$$

Or, α étant très-petit, l'accroissement $\tang(C+\alpha) - \tang C$ est sensiblement égal (corollaire II) au produit de α par la dérivée de $\tang C$, c'est-à-dire à $\frac{\alpha}{\cos^2 C}$. On aura donc

$$\frac{\varepsilon}{h} = \frac{\alpha}{\sin C \cos C} = \frac{2\alpha}{\sin 2C}.$$

Or c'est un résultat de l'expérience que l'erreur α est indépendante de la grandeur de l'angle C; mais on voit que l'erreur relative $\frac{\varepsilon}{h}$ change, pour une même valeur de α, avec la grandeur de l'angle C. Pour qu'elle soit minimum, il faut que $\sin 2C$ soit maximum ou que C égale 45°. Ainsi le triangle rectangle le plus *avantageux*, c'est-à-dire celui pour lequel une erreur constante, commise sur l'angle observé, influera le moins possible sur la hauteur cherchée, est un triangle isocèle. C'est pourquoi, dans la pratique, on a soin de choisir une base b qui diffère peu de la hauteur, en sorte que l'angle C se rapproche le plus possible de 45°.

La *règle des parties proportionnelles*, employée dans les calculs logarithmiques, consiste à admettre que les différences entre les nombres (ou arcs) sont proportion-

nelles aux différences entre les logarithmes des nombres (ou lignes trigonométriques des arcs).

Ainsi, soit proposé de calculer le logarithme du sinus de l'arc $35^\circ\, 8'\, 47'',5$. Les Tables donnent

$$\log \sin (35^\circ\, 8'\, 40'') = \bar{1},7601508\,;$$

et la différence tabulaire correspondante, c'est-à-dire l'excès de $\log \sin (35^\circ\, 8'\, 50'')$ sur le logarithme précédent, est $299^{\text{dix-million.}}$. Si l'on désigne par δ le nombre de dix millioniémes qu'il faut ajouter à $\log \sin (35^\circ\, 8'\, 40'')$ pour avoir le logarithme demandé, on fait la proportion,

$$\frac{\delta}{299} = \frac{7,5}{10}, \quad \text{d'où} \quad \delta = 224,$$

et par suite

$$\log \sin (35^\circ\, 8'\, 47'',5) = \bar{1},7601732.$$

Nous déterminerons (63) une limite supérieure de l'erreur que cette règle comporte.

54. La démonstration donnée au corollaire II du n° 52 exige que la dérivée $f'(x)$ ne soit pas nulle pour la valeur particulière $x = a$. Si l'on avait $f'(a) = 0$, le terme $hf'(a)$ disparaîtrait, et il ne serait plus vrai de dire que l'accroissement $f(a+h) - f(a)$ est sensiblement proportionnel à h. Cet accroissement serait alors comparable à une puissance de h supérieure à la première. C'est ce que l'on verra par la formule de Taylor (65). Ce cas exceptionnel se présentera notamment, quand la fonction $f(x)$ passera, pour la valeur $x = a$, par un maximum ou un minimum (49).

Soit, par exemple, $f(x) = \cos x$: partons de la valeur $a = 0^\circ$, qui rend $\cos x$ maximum. Il vient

$$f(0^\circ + h) - f(0^\circ) = \cos h - 1 = -2 \sin^2 \frac{h}{2} = -\frac{h^2}{2} \left(\frac{\sin \frac{h}{2}}{\frac{h}{2}} \right)^2.$$

Pour de très-petites valeurs de h, le facteur $\frac{\sin \frac{h}{2}}{\frac{h}{2}}$ diffère très-peu de 1, et par suite l'accroissement de la fonction $\cos x$, en partant de $x = 0°$, est sensiblement proportionnel à h^2. Il en résulte que cette fonction varie très-lentement dans le voisinage de son maximum. Ce fait a lieu pour toute fonction dont la dérivée est continue.

55. La proportionnalité de l'accroissement d'une fonction à l'accroissement de la variable dont elle dépend, n'aura lieu rigoureusement qu'autant que $f'(a + \theta h)$ sera indépendante de h, ou, ce qui revient au même, que $f'(x)$ sera une constante. Soit A cette constante. Comme $f'(x) - \mathrm{A}$ est la dérivée de $f(x) - \mathrm{A}x$, il suit de l'équation $f'(x) - \mathrm{A} = 0$, que la fonction $f(x) - \mathrm{A}x$ sera elle-même une constante (49 et 52). Soit B cette seconde constante. On aura donc

$$f(x) = \mathrm{A}x + \mathrm{B}.$$

Ainsi, pour que l'accroissement d'une fonction de x soit constamment proportionnel à l'accroissement de la variable, il faut et il suffit que cette fonction soit entière et du premier degré en x.

L'interprétation géométrique de ce résultat est trop évidente pour que nous nous y arrêtions.

56. *La règle de fausse position*, ou *des parties proportionnelles*, sert à résoudre approximativement une équation de degré quelconque à l'aide de deux suppositions successives faites sur la valeur de l'inconnue.

Soient M et N deux grandeurs (deux temps, deux distances, deux volumes, etc.) qui dépendent d'une inconnue x, qu'il s'agit de déterminer de manière que ces grandeurs deviennent égales. Si l'on substitue dans l'équa

tion

$$M - N = o,$$

à la place de x, une valeur a arbitraire, mais déjà approchée de la racine cherchée, M ne sera pas égale à N, et la différence $M - N$ prendra une certaine valeur A. Pour une seconde supposition $a + h$ faite sur l'inconnue, $M - N$ prendra une autre valeur que nous désignerons par $A + \Delta A$; ΔA sera donc l'accroissement (positif ou négatif) que prend la fonction $M - N$ pour l'accroissement h donné à la variable. Cela posé, en admettant la proportionnalité, on dira : Si pour un accroissement h donné à x, la fonction prend l'accroissement ΔA, quel sera l'accroissement X à donner à x, pour que la fonction prenne l'accroissement $-A$, c'est-à-dire se réduise à zéro? d'où la proportion

$$\frac{X}{h} = \frac{-A}{\Delta A}, \quad \text{et} \quad X = -\frac{Ah}{\Delta A}.$$

Dans le cas où cette proportion sera exacte, $a + X$, c'est-à-dire $a - \frac{Ah}{\Delta A}$, sera la valeur de l'inconnue qui vérifie l'équation $M - N = o$; c'est ce qui aura lieu si la fonction $M - N$ est du premier degré en x. Dans le cas contraire, la proportion ne fournira qu'une approximation plus ou moins grande, selon le degré de petitesse de h et de A.

Nous donnerons de nombreuses applications de cette règle dans le calcul approché des racines des équations numériques (n°s 79 et suivants).

57. Quel que soit le calcul à faire sur une grandeur a, on peut représenter par $f(a)$ le résultat cherché; $f(a - \alpha)$ sera le résultat correspondant à la valeur approchée $a - \alpha$. L'erreur commise, que nous désignerons par ε, sera la valeur absolue de la différence $f(a) - f(a - \alpha)$, et le

signe de cette différence apprendra si le résultat du calcul est approché par défaut ou par excès.

Or, si l'on remplace, dans l'équation (1), h par α et a par $a-\alpha$, il vient

$$f(a)-f(a-\alpha)=\alpha f'(a-\theta\alpha).$$

θ n'est plus le même que dans l'équation (1); mais c'est toujours un nombre compris entre 0 et 1, en sorte que $a-\theta\alpha$ est encore une moyenne entre a et la valeur $a-\alpha$. On a donc, en prenant la valeur absolue du second membre,

$$(2) \qquad \varepsilon=\alpha f'(a-\theta\alpha).$$

Si l'on opérait sur un nombre en excès, $a+\alpha$, on trouverait de même pour l'expression numérique de l'erreur,

$$(2\ bis.) \qquad \varepsilon'=\alpha f'(a+\theta\alpha);$$

$a+\theta\alpha$ est encore une valeur moyenne entre a et la valeur $a+\alpha$ employée dans le calcul.

Ainsi, *l'erreur absolue commise dans un calcul quelconque fait sur une grandeur* a *dont on n'a qu'une valeur approchée* a $\pm\alpha$, *est égale au produit de* α *par la valeur que prend la dérivée de la fonction qui représente le résultat du calcul à effectuer, lorsqu'on y remplace* a *par une moyenne entre* a *et le nombre employé dans le calcul.*

A la vérité, cette valeur moyenne n'est pas connue; mais on saura, dans chaque cas particulier, par une substitution convenable, tirer de la formule précédente une limite supérieure de l'erreur commise, et c'est cette limite seule qu'il importe d'avoir. Donnons des exemples.

58. Soit d'abord

$$f(x)=x^m, \quad \text{d'où} \quad f'(x)=mx^{m-1};$$

on aura, par la formule (2)

$$\varepsilon=m(a-\theta\alpha)^{m-1}\alpha.$$

Cette formule convient à toute valeur de m. Si m est plus grand que 1, on en tire

$$\varepsilon < m a^{m-1} \alpha,$$

ce qui est conforme au résultat trouvé (35).

Si m est plus petit que 1, soit $m = \frac{1}{n}$; $n > 1$. L'équation précédente devient

$$\varepsilon = \frac{\alpha}{n \sqrt[n]{(a - \theta\alpha)^{n-1}}};$$

d'où l'on tire

$$\varepsilon < \frac{\alpha}{n \sqrt[n]{(a - \alpha)^{n-1}}},$$

comme on l'a vu (37).

Exemple. Soit proposé de calculer $\sqrt[3]{\frac{375300}{\pi}}$ à une unité près. Posons

$$f(x) = \sqrt[3]{\frac{375300}{x}}, \quad \text{d'où} \quad f'(x) = -\frac{\sqrt[3]{375300}}{3x\sqrt[3]{x}}.$$

L'erreur ε, commise en prenant pour x, $\pi + \alpha$ au lieu de π, sera, par la formule (2 *bis*),

$$\varepsilon = \frac{\alpha \sqrt[3]{375300}}{3(\pi + \theta\alpha)\sqrt[3]{\pi + \theta\alpha}} < \frac{\alpha \sqrt[3]{375300}}{3\pi \sqrt[3]{\pi}},$$

et, à fortiori, l'on aura

$$\varepsilon < \frac{80\,a}{9} < 10\,\alpha;$$

or la question demande que ε soit moindre que 1 : il suffit donc de poser

$$\alpha < \frac{1}{10}.$$

Ainsi l'on prendra pour π la valeur par excès 3,2, et il restera à extraire la racine cubique du quotient $\frac{375300}{3,2}$ à une unité près. A cet effet on calculera seulement les trois premiers chiffres 117 de ce quotient (38). La racine cubique de 117000 est 48, et par conséquent 49 est la valeur approchée de $\sqrt[3]{\frac{375300}{\pi}}$ à une unité près.

Cette méthode s'appliquera très-bien à toutes les opérations complexes du même genre, dans lesquelles on n'aura à considérer qu'*une seule* grandeur incommensurable. Son principal avantage est de *déterminer l'approximation par une seule formule*, tandis qu'autrement il faut successivement recourir, dans une même question, à plusieurs formules distinctes et propres à chaque espèce d'opérations. C'est ce qu'on a fait au n° 39. Au surplus, l'emploi des Tables de logarithmes dans les questions de cette nature fournira beaucoup plus rapidement l'approximation demandée.

59. La formule générale que nous venons d'exposer s'applique sans plus de difficulté au cas où la fonction $f(x)$ est transcendante.

Ainsi, soit

$$f(x) = \log x, \quad \text{d'où} \quad f'(x) = \frac{\log e}{x},$$

e désignant la base des logarithmes népériens. Si le logarithme que nous considérons est pris dans la base 10, le module $\log e$ ou $\frac{1}{l\,10}$ est égal (30) à 0,43429..., et, par conséquent, moindre que $\frac{1}{2}$.

Or l'erreur ε que l'on commettra sur $\log a$, en prenant, au lieu de a, la valeur approchée $a - \alpha$, sera, d'après la formule (2),

$$\varepsilon = \frac{\alpha \log e}{a - \theta\alpha},$$

et, puisque l'on a $\log e < \frac{1}{2}$ et $\theta < 1$, il viendra

$$\varepsilon < \frac{1}{2} \cdot \frac{\alpha}{a - \alpha}.$$

Supposons, par exemple, $a - \alpha = 345,7$ et $\alpha < \frac{1}{10}$, on aura

$$\varepsilon < \frac{1}{20 . 345,7} \quad \text{ou} \quad \varepsilon < \frac{1}{1000};$$

ainsi l'on ne pourra compter que sur les trois premières décimales du logarithme du nombre qui a pour valeur approchée 345,7 à $\frac{1}{10}$ près. Il serait donc illusoire, en pareil cas, de faire usage des sept décimales que donnent les Tables de Callet. Si l'on prend, dans ces Tables, les logarithmes de deux nombres consécutifs 345,7 et 345,8, entre lesquels tombe le nombre a dont il s'agit, on trouve

$$\log 345,7 = 2,5386994,$$
$$\log 345,8 = 2,5388250,$$

et l'on voit, en effet, que ces deux logarithmes n'ont que les trois premières décimales communes.

60. Pour que l'erreur ε qui affecte $\log(a - \alpha)$ soit moindre qu'une unité du septième ordre décimal, il faudra que α satisfasse à la condition

$$\frac{1}{2} \cdot \frac{\alpha}{a - \alpha} < \frac{1}{10^7},$$

d'où

$$\frac{\alpha}{a} < \frac{2}{10^7 + 2};$$

Cette inégalité sera évidemment satisfaite si l'on a

$$\frac{\alpha}{a} < \frac{1}{10^7}.$$

Donc (5) le nombre a devra être connu avec huit chiffres

exacts. *En général, il faut connaître un chiffre de plus dans le nombre* a *qu'on n'en veut avoir dans son logarithme.* Par exemple, soit proposé de calculer, à l'aide des Tables de Callet, le logarithme du nombre π avec sept chiffres. On prendra π avec huit chiffres, ou $\pi = 3{,}1415926$, et l'on trouvera

$$\log\pi = 0{,}49714987\ldots.$$

Comme on ne doit conserver que sept décimales, il convient (8) d'augmenter de 1 la dernière, et l'on a définitivement,

$$\log\pi = 0{,}4971499.$$

Ce résultat est même approché à moins d'une demi-unité du dernier ordre.

Nous n'avons pas égard ici à l'erreur qui provient de la règle *des parties proportionnelles,* attendu qu'elle n'influe pas, comme on va le voir, sur les sept premières décimales du logarithme : la discussion de cette erreur fera l'objet des paragraphes suivants.

61. Supposons d'abord qu'il s'agisse de *calculer le logarithme d'un nombre donné.*

Comme la question se ramène toujours à chercher le logarithme d'un nombre compris entre deux termes consécutifs de la Table, nous désignerons ce nombre par $a+h$, a étant l'entier immédiatement inférieur, dont le logarithme se lit dans la Table. En posant

$$f(x) = \log x,$$

on a

$$f'(x) = \frac{\log e}{x};$$

et, d'après l'équation (1) du n° 50,

$$\log(a+h) - \log a = \frac{h\log e}{a+\theta h}.$$

Pour $h=1$, il vient

$$\log(a+1) - \log a = \frac{\log e}{a+\theta_1};$$

θ a changé, parce qu'il dépend de h; c'est pourquoi nous mettons θ_1. Du reste θ_1, comme θ, est toujours compris entre o et 1.

Désignons, pour abréger, les premiers membres des deux égalités précédentes par δ et Δ; et retranchons de la première le produit de la seconde par h, il viendra

$$(3) \qquad \delta - h\Delta = h \log e \frac{\theta_1 - \theta h}{(a + \theta h)(a + \theta_1)}.$$

La valeur numérique du second membre de cette égalité n'est autre que l'erreur que l'on commet sur δ, lorsqu'on prend, d'après la règle des parties proportionnelles (53), $\delta = h\Delta$. Or on a

$$\log e < \frac{1}{2}, \quad h < 1, \quad \theta_1 - \theta h < 1.$$

Donc la valeur numérique de $\delta - h\Delta$ est moindre que $\frac{1}{2a^2}$. Dans les Tables de Callet, on ne fait usage des parties proportionnelles que pour les nombres supérieurs à 10000; l'erreur $\delta - h\Delta$ est alors moindre que $\frac{1}{2.10^8}$; et, par conséquent, elle n'atteint même pas la huitième décimale. Toutefois à cette erreur s'ajoute celle qui résulte de ce que le facteur Δ du produit $h\Delta$ n'est approché qu'à une demi-unité du septième ordre, et cette dernière a pour limite supérieure $\frac{1}{2.10^7}$. Les deux erreurs réunies donnent toujours une somme moindre que $\frac{1}{10^7}$. On peut donc compter sur les sept premières décimales du logarithme.

62. Dans le problème inverse, c'est-à-dire lorsqu'on cherche le *nombre correspondant à un logarithme donné*, h est la fraction inconnue qu'il faut ajouter à l'entier a pour avoir le nombre cherché. Or l'équation (3) peut s'é-

crire

$$\delta - h\Delta = h\Delta \frac{\theta_1 - \theta h}{a + \theta h},$$

d'où

$$\frac{\delta}{\Delta} - h = \frac{h(\theta_1 - \theta h)}{a + \theta h}.$$

L'erreur commise en prenant pour h le quotient $\frac{\delta}{\Delta}$ n'est autre que la valeur numérique du second membre de cette égalité. Or on a évidemment

$$\frac{h(\theta_1 - \theta h)}{a + \theta h} < \frac{1}{a}.$$

Donc, si l'on n'applique la règle des parties proportionnelles qu'aux nombres plus grands que 10000, l'erreur commise en remontant des logarithmes aux nombres sera moindre que $\frac{1}{10000}$. Mais il ne faut pas en conclure qu'on puisse compter sur les quatre premiers chiffres décimaux du quotient $\frac{\delta}{\Delta}$, attendu qu'à l'erreur précédente s'ajoute celle qui provient de ce que le dividende δ et le diviseur Δ ne sont eux-mêmes approchés qu'à une demi-unité près. Cette dernière a pour limite supérieure $\frac{1}{\Delta}$, comme on l'a vu au nº 27. Or, dans les Tables de Callet, la différence tabulaire Δ est un nombre de *deux* ou *trois* chiffres; c'est pourquoi l'on recommande en général de ne pas pousser le calcul de h au delà du chiffre des centièmes.

63. Proposons-nous encore d'apprécier l'erreur que la règle des parties proportionnelles comporte dans le calcul des logarithmes-sinus.

En posant

$$f(x) = \log \sin x,$$

on a

$$f'(x) = \frac{\log e}{\tang x}.$$

x désigne la longueur de l'arc qui mesure un angle quelconque dans le cercle décrit du sommet de cet angle comme centre avec un rayon égal à 1 : soient a le nombre de dizaines de secondes immédiatement inférieur à l'arc dont on cherche le logarithme-sinus, et h le nombre (entier ou fractionnaire) de secondes dont cet arc surpasse a''. h est donné; la Table donne log sin a'' ainsi que log sin $(a + 10'')$, et il s'agit de calculer log sin $(a + h'')$. Avant d'aller plus loin, il convient de remarquer que, tant que a et h resteront *engagés sous les signes trigonométriques, sin, tang*, il n'y aura aucun inconvénient à supprimer le double accent $''$ destiné à rappeler que ces lettres a et h désignent des arcs exprimés en secondes, et à écrire log sin a, log sin $(a + h)$... : c'est ce que l'on fait d'ordinaire. Mais si, par suite du calcul, l'une de ces lettres, h par exemple, vient à *sortir des signes trigonométriques*, il ne faudra pas oublier que cette lettre désigne l'arc de h'' ou h fois la longueur de l'arc $1''$.

Cela posé, on a, d'après l'équation (1) du n° 50,

$$\log\sin(a+h) - \log\sin a \quad \text{ou} \quad \delta = \frac{h\log e}{\tang(a+\theta h)}\,\text{arc}\,1'',$$

$$\log\sin(a+10'') - \log\sin a \quad \text{ou} \quad \Delta = \frac{10\log e}{\tang(a+\theta_1 . 10'')}\,\text{arc}\,1''.$$

Retranchons de la première égalité le produit de la seconde par $\frac{h}{10}$; il vient,

$$\delta - \frac{h}{10}\Delta = h\log e.\text{arc}\,1''\left[\frac{1}{\tang(a+\theta h)} - \frac{1}{\tang(a+\theta_1 . 10'')}\right],$$

ou

$$(4)\quad \delta - \frac{h}{10}\Delta = h\log e.\text{arc}\,1'' \cdot \frac{\sin(\theta_1 . 10'' - \theta h)}{\sin(a+\theta h)\sin(a+\theta_1 . 10'')}.$$

La valeur numérique du second nombre est l'erreur commise sur δ, lorsqu'on prend, conformément à la règle

usuelle, $\delta = \frac{h}{10}\Delta$. Or on a

$$\log e < \frac{1}{2}, \quad h < 10, \quad \sin(\theta_1 10'' - \theta h) < \sin 10'' < \operatorname{arc} 10'',$$

et

$$\sin(a + \theta h)\sin(a + \theta_1 . 10'') > \sin^2 a;$$

donc l'erreur commise est moindre que

$$\frac{(\operatorname{arc} 10'')^2}{2\sin^2 a}.$$

D'ailleurs, on a

$$\operatorname{arc} 10'' = \frac{\pi}{64800} = 0,000048\ldots < \frac{1}{2.10^4};$$

donc enfin l'erreur commise par la proportion des Tables a pour limite supérieure

$$\frac{1}{8.10^8 \sin^2 a}.$$

Si l'on part d'un arc $a = 5^\circ$, on trouve que cette limite se réduit sensiblement à $\frac{1}{10^7.0,6}$. Ainsi l'on pourrait craindre que la septième décimale ne fût fautive d'*une unité*. Mais cela n'a même pas lieu. On verra en effet (73) que la limite d'erreur trouvée ci-dessus peut être réduite au quart de sa valeur, ou à

$$\frac{1}{32.10^8.\sin^2 a}.$$

Si l'on adopte ce dernier nombre, l'erreur effective, pour un arc $a = 5^\circ$, et à fortiori pour un arc plus grand, a pour limite supérieure $\frac{1}{10^7.2,4}$ ou une demi-unité du septième ordre.

Mais pour un arc compris entre 0 et 5 degrés, et surtout s'il est voisin de 0 degré, on ne saurait répondre que l'er-

reur provenant de la proportion des Tables n'atteindra pas la septième décimale. C'est pourquoi l'on a construit une Table spéciale renfermant les logarithmes-sinus et logarithmes-tangentes, *de seconde en seconde*, *pour les cinq premiers degrés*.

64. Dans le problème inverse, on cherche l'arc $a+h$ correspondant à un logarithme-sinus donné ; c'est le nombre de secondes désigné par h qui est l'inconnue. Or l'équation (4) peut s'écrire

$$\delta - \frac{h}{10}\Delta = \frac{h}{10}\Delta \frac{\sin(\theta_1 . 10'' - \theta h)}{\sin(a+\theta h)\cos(a+\theta_1 . 10'')},$$

d'où

$$\frac{\delta}{\Delta} - \frac{h}{10} = \frac{h}{10} . \frac{\sin(\theta_1 . 10'' - \theta h)}{\sin(a+\theta h)\cos(a+\theta_1 . 10'')}.$$

L'erreur que l'on fait sur la fraction $\frac{h}{10}$ en la prenant égale à $\frac{\delta}{\Delta}$, est donc la valeur numérique du rapport

$$\frac{h}{10} . \frac{\sin(\theta_1 . 10'' - \theta h)}{\sin(a+\theta h)\cos(a+\theta_1 . 10'')},$$

et par conséquent moindre que

$$\frac{\text{arc } 10''}{\sin a \cos(a+10'')} \quad \text{ou à très-peu près} \quad \frac{1}{10^4 . \sin 2a}.$$

Si l'on n'applique la proportion des Tables qu'à des arcs égaux ou supérieurs à 5 degrés, on aura :

$$\sin 2a \geqq \sin 10^\circ = 0,17\ldots,$$

et par suite l'erreur sera moindre que $\frac{1}{10^3}$. Elle n'affectera donc pas le chiffre des millièmes du rapport $\frac{h}{10}$; ou, ce qui revient au même, le chiffre des centièmes de h. Mais il ne faut pas oublier qu'à cette erreur s'en joint une autre, pro-

venant de ce que les deux termes du quotient $\frac{\delta}{\Delta}$ ne sont connus qu'à une demi-unité près, erreur dont la limite supérieure est, comme on sait, $\frac{1}{\Delta}$; par ce fait seul, le quotient $\frac{10\delta}{\Delta}$ n'est approché qu'à $\frac{1}{\left(\frac{\Delta}{10}\right)}$ près. Cette seconde erreur peut atteindre les centièmes de seconde, attendu que la différence tabulaire Δ est un nombre de trois ou de quatre chiffres au plus (*). C'est pourquoi l'on s'arrête dans le calcul de h au chiffre des centièmes, et même on ne conserve communément que celui des dixièmes.

65. Le calcul approché des logarithmes par la règle des parties proportionnelles est un cas particulier du problème de l'*interpolation*. Interpoler, c'est déterminer, entre certaines limites de la variable x, les valeurs d'une fonction de x, d'après la connaissance d'un certain nombre de valeurs particulières de cette fonction, correspondant à des valeurs de x comprises entre les mêmes limites. Ici les valeurs connues de la fonction $\log x$ ou $\log \sin x$ sont les termes consécutifs des Tables, lesquels répondent à des valeurs équidistantes de x. Or, si l'on ouvre les Tables des nombres ou des lignes trigonométriques, on observe que les différences entre chaque terme et le suivant, ou *différences premières*, varient très-lentement, en sorte qu'on peut les regarder comme constantes dans un certain intervalle, c'est-à-dire regarder comme nulles les différences de ces différences, ou *différences secondes*, et celles des ordres supérieurs. Il en résulte que les fonctions $\log x$, $\log \sin x$, $\log \operatorname{tang} x$, etc., se confondent sensiblement, pour les valeurs de x comprises entre deux

(*) Il est de quatre chiffres pour les arcs inférieurs à 11° 54′, et de trois chiffres pour tous les autres jusqu'à 45 degrés.

termes consécutifs de la Table, avec des fonctions entières et du premier degré en x; et l'on sait (55) que, pour ces dernières, la proportionnalité entre les accroissements de la fonction et ceux de la variable a lieu rigoureusement. Le lecteur pourra consulter, sur ce point détaché de la théorie des différences et sur l'application au calcul approché des racines des équations, l'article inséré dans les *Nouvelles Annales de Mathématiques* (janvier et février 1851).

66. L'équation (1) nous a fourni l'expression générale de l'erreur commise quand on réduit $f(a+h)$ à $f(a)$, c'est-à-dire quand on néglige dans le calcul de $f(a+h)$ tout ce qui dépend de h. L'erreur $[hf'(a+\theta h)]$ est, en général, du même ordre que h; son rapport à h tend vers une limite finie, lorsque h tend vers zéro. Cette substitution de $f(a)$ à $f(a+h)$ constitue un *premier ordre d'approximation*; mais il arrive souvent que l'on a besoin d'une approximation plus grande. On complète alors le terme $f(a)$ par un second terme proportionnel à h, et l'erreur n'est plus que de l'ordre du carré de h; c'est le *second ordre d'approximation*. C'est ainsi que par les formules des nos 32, 44 et 46, nous avons remplacé

$$\frac{1}{1+x} \text{ par } 1-x, \text{ à moins de } x^2,$$

$$\sqrt{1+\varepsilon} \text{ par } 1+\frac{\varepsilon}{2}, \text{ à moins de } \frac{\varepsilon^2}{8},$$

$$\sqrt[3]{1+\varepsilon} \text{ par } 1+\frac{\varepsilon}{3}, \text{ à moins de } \frac{\varepsilon^2}{8}.$$

Les quantités x, ε, que l'on suppose très-petites, jouent le rôle de h dans $f(a+h)$.

On conçoit de même un *troisième ordre d'approximation*, dans lequel on négligerait seulement les termes du développement de $f(a+h)$ qui renferment le cube et les

puissances supérieures de h; mais, comme h est supposé très-petit relativement à a, il n'arrive guère dans la pratique que l'on dépasse le second ordre d'approximation.

67. Le problème général de l'approximation, pour tous les ordres, serait résolu, si l'on connaissait le développement d'une fonction quelconque $f(a+h)$ suivant les puissances entières et positives de h, et le reste qui complète la somme des n premiers termes de ce développement. Tel est l'objet de la formule de Taylor.

Quand il s'agit d'une fonction $f(x)$ entière et du degré m, la formule de Taylor s'établit aisément par la loi du binôme de Newton, et l'on trouve

$$f(a+h)=f(a)+hf'(a)+\frac{h^2}{1.2}f''(a)+\ldots+\frac{h^m}{1.2.3\ldots m}f^m(a).$$

Ce développement se termine au $m+1^{ième}$ terme; car les dérivées, $f^{m+1}(a)$, $f^{m+2}(a)$ etc., sont toutes nulles.

Si la fonction $f(x)$ n'est plus entière, mais quelconque, on peut néanmoins se proposer de développer encore $f(a+h)$ suivant la loi précédente; mais le développement va présenter une *série* illimitée de termes. Que si l'on s'arrête au $n^{ième}$ terme et qu'on désigne par R le *reste* inconnu destiné à compléter la série, on aura par définition,

$$(5)\qquad \left\{\begin{array}{l} R=f(a+h)-f(a)-hf'(a)-\frac{h^2}{1.2}f''(a)-\ldots \\ \quad -\frac{h^{n-1}}{1.2.3\ldots(n-1)}f^{n-1}(a), \end{array}\right.$$

et la question consistera à trouver une expression de ce reste sous forme finie. Voici cette expression :

$$(6)\qquad R=\frac{h^n}{1.2.3\ldots n}f^n(a+\theta h),\quad [0<\theta<1].$$

Elle est applicable à toute fonction $f(x)$ algébrique ou transcendante, pourvu que la $n^{ième}$ dérivée $f^n(x)$ reste finie

et continue pour toute valeur de x comprise entre a et $a+h$. Comme cas particulier, si l'on y fait $n=1$, il vient

$$R=\frac{h}{1}f'(a+\theta h),$$

et par suite, en vertu de l'équation (5), où l'on doit faire en même temps $n=1$,

$$f(a+h)-f(a)=hf'(a+\theta h);$$

c'est la formule (1) du n° 50.

Pour $n=2$, l'expression (6) devient

$$R=\frac{h^2}{1.2}f''(a+\theta h)$$

(θ a changé, car il dépend de n; mais il reste toujours compris entre o et 1); et, d'après l'équation (5), où l'on fait $n=2$, on a

$$(7)\quad f(a+h)-f(a)-hf'(a)=\frac{h^2}{1.2}f''(a+\theta h).$$

Cette dernière équation n'est pas moins utile que l'équation (1) dans la théorie des approximations numériques : elle est au second ordre de l'approximation ce que celle-ci est au premier ordre.

68. Comme nous n'aurons pas besoin ici de la formule de Taylor dans toute sa généralité, il suffira de démontrer la formule (7). Au reste, le lecteur curieux de connaître la démonstration générale la trouvera en petits caractères dans le numéro suivant.

Supposons d'abord h positif, et posons

$$\varphi(x)=f(a+h)-f(x)-(a+h-x)f'(x).$$

Pour $x=a$, $\varphi(x)$ devient $f(a+h)-f(a)-hf'(a)$,

et se confond avec le reste R dont nous cherchons l'expression. En prenant la dérivée de $\varphi(x)$, il vient

$$\varphi'(x) = -(a+h-x)f''(x).$$

Soit M la plus grande des valeurs que prend $f''(x)$, lorsque x croît d'une manière continue de a à $a+h$, nous aurons constamment

$$(A) \qquad \frac{a+h-x}{1}[M - f''(x)] \geqq 0;$$

donc la fonction qui a pour dérivée le premier membre de cette inégalité est croissante. Or cette fonction est, d'après ce qui précède,

$$\varphi(x) - \frac{(a+h-x)^2}{1.2}M;$$

ainsi cette fonction est croissante, et comme elle finit par s'annuler pour $x = a+h$, elle a été constamment négative entre a et $a+h$. Si l'on y fait $x = a$, on aura donc

$$\varphi(a) \text{ ou } R < \frac{h^2}{1.2}M.$$

En désignant par m la plus petite des valeurs que prend $f''(x)$ entre a et $a+h$, on démontrera de même l'inégalité

$$R > \frac{h^2}{1.2}m.$$

La valeur de R étant comprise entre $\frac{h^2}{1.2}m$ et $\frac{h^2}{1.2}M$, on peut l'égaler au produit de $\frac{h^2}{1.2}$ par une certaine grandeur moyenne entre m et M.

Nous avons supposé h positif. Dans le cas contraire, on fera croître x depuis $a+h$ jusqu'à a; le facteur $a+h-x$ sera négatif, et l'inégalité (A) sera renversée. La

fonction $\varphi(x) - \frac{(a+h-x)^2}{1.2}$ M sera donc décroissante, et comme elle *commence* par s'annuler pour $x = a + h$, elle sera constamment négative entre $a + h$ et a; donc on aura encore pour $x = a$,

$$R < \frac{h^2}{1.2} M.$$

Sans qu'il soit nécessaire de poursuivre, on voit que les conclusions restent les mêmes.

Tout ceci n'exige pas que la dérivée seconde soit *continue*, mais seulement *finie*, entre a et $a + h$. Maintenant, si l'on suppose $f''(x)$ continue, elle passera par tous les états de grandeur entre sa plus grande et sa plus petite valeur; par suite, il existera une valeur $a + \theta h$ de x, comprise entre a et $a + h$, pour laquelle $f''(x)$ atteindra cette grandeur moyenne entre m et M, dont il a été question plus haut, et dont le produit par $\frac{h^2}{1.2}$ est égal au reste R. On aura donc

$$R = \frac{h^2}{1.2} f''(a + \theta h),$$

et la formule (7) est démontrée.

69. En général, soit

$$\varphi(x) = f(a+h) - f(x) - (a+h-x)f'(x)$$
$$- \frac{(a+h-x)^2}{1.2} f''(x) - \ldots - \frac{(a+h-x)^{n-1}}{1.2.3\ldots(n-1)} f^{n-1}(x).$$

Pour $x = a$, $\varphi(a)$ se confond avec l'expression (5) du reste R.

La dérivée de $\varphi(x)$ se réduit à

$$\varphi'(x) = - \frac{(a+h-x)^{n-1}}{1.2.3\ldots(n-1)} f^n(x).$$

On suppose d'abord h positif, et si l'on désigne par M la plus grande des valeurs que prend $f^n(x)$ lorsque x croît de a à $a + h$,

on a constamment

$$\text{(B)} \qquad \frac{(a+h-x)^{n-1}}{1.2.3\ldots(n-1)}[\mathrm{M}-f^n(x)] > 0;$$

donc la fonction qui a pour dérivée le premier membre de cette inégalité est croissante. Or cette fonction est évidemment

$$\varphi(x) - \frac{(a+h-x)^n}{1.2.3\ldots n}\mathrm{M};$$

et comme elle finit par s'annuler pour $x = a + h$, elle a dû être constamment négative pour les valeurs inférieures de x. On a donc en particulier pour $x = a$,

$$\varphi(a) - \frac{h^n}{1.2.3\ldots n}\mathrm{M} < 0, \quad \text{ou bien} \quad \mathrm{R} < \frac{h^n}{1.2.3\ldots n}\mathrm{M}.$$

En désignant par m la plus petite des valeurs que prend $f^n(x)$, lorsque x varie de a à $a + h$, on aura de même

$$\mathrm{R} > \frac{h^n}{1.2.3\ldots n}m.$$

Ainsi R est égal au produit de $\frac{h^n}{1.2.3\ldots n}$ par une certaine quantité moyenne entre m et M, qu'on peut représenter par $f^n(a+\theta h)$, si $f^n(x)$ est continue entre a et $a + h$. On a donc

$$(6) \qquad \mathrm{R} = \frac{h^n}{1.2.3\ldots n}f^n(a+\theta h).$$

Nous avons supposé h positif. Dans le cas contraire, on fera croître x depuis $a + h$ jusqu'à a; le facteur $a + h - x$ sera négatif, et l'on aura à distinguer deux cas : n pair, n impair.

1°. n pair; $(a+h-x)^{n-1}$ étant négatif, l'inégalité (B) change de signe. Donc la fonction

$$\varphi(x) - \frac{(a+h-x)^n}{1.2.3\ldots n}\mathrm{M}$$

est décroissante quand x croît de $a + h$ à a, et comme elle com-

mence par s'annuler pour $x = a + h$, elle doit être constamment négative pour les valeurs supérieures de x. On a donc en particulier, pour $x = a$,

$$\varphi(a) \quad \text{ou} \quad R < \frac{h^n}{1.2.3\ldots n} M,$$

comme précédemment, etc. Les conclusions restent les mêmes.

2°. n impair; $(a + h - x)^{n-1}$ est positif, et l'inégalité (B) conserve son signe. La fonction

$$\varphi(x) - \frac{(a + h - x)^n}{1.2.3\ldots n} M$$

est donc croissante, et, par suite, positive pour les valeurs de x supérieures à $a + h$. On a donc, pour $x = a$,

$$\varphi(a) \quad \text{ou} \quad R > \frac{h^n}{1.2.3\ldots n} M.$$

On démontrerait de même que l'on a

$$R < \frac{h^n}{1.2.3\ldots n} m.$$

Ces deux inégalités n'ont rien de contradictoire, vu que h^n est négatif.

En définitive, R est encore égal au produit de $\frac{h^n}{1.2.3\ldots n}$ par une moyenne entre m et M. La formule (6) subsiste donc dans tous les cas. Par suite, eu égard à l'équation (5), on a

$$f(a + h) = f(a) + hf'(a) + \frac{h^2}{1.2} f''(a) + \ldots$$
$$+ \frac{h^{n-1}}{1.2.3\ldots(n-1)} f^{n-1}(a) + \frac{h^n}{1.2.3\ldots n} f^n(a + \theta h).$$

C'est la formule de Taylor.

70. Cette formule s'étend aux fonctions de plusieurs variables indépendantes. Considérons, par exemple, une fonction de deux variables $f(x, y)$, dans laquelle on don-

nera successivement aux variables les deux couples de valeurs

$$(x = a,\ y = b),\quad (x = a + h,\ y = b + k).$$

On peut se proposer de développer $f(a + h, b + k)$ suivant les puissances de h et de k. Ce développement (que nous limiterons, pour abréger, aux termes affectés des premières puissances de h et de k) est le suivant :

$$(8)\quad f(a + h, b + k) = f(a, b) + \left[h f'_a(a, b) + k f'_b(a, b)\right] + \mathrm{R},$$

le reste R ayant pour expression

$$\mathrm{R} = \frac{1}{1.2}\left[\begin{array}{c} h^2 f''_{a,a}(a + \theta h, b + \theta k) + 2hk f''_{a,b}(a + \theta h, b + \theta k) \\ + k^2 f''_{b,b}(a + \theta h, b + \theta k) \end{array}\right];$$

$$0 < \theta < 1.$$

(La notation f'_a indique qu'on a pris la dérivée partielle de $f(a, b)$ par rapport à a ; $f''_{a,b}$ indique qu'on a pris ensuite la dérivée par rapport à b de cette première dérivée partielle ; $f''_{a,a}$ indique la dérivée par rapport à a de la première dérivée relative à cette même variable, etc.) Nous nous bornons à énoncer cette formule qu'on peut déduire de celle qui vient d'être démontrée pour une seule variable.

71. Si l'on désigne par h et k les erreurs commises sur deux grandeurs dont a et b sont les valeurs employées dans le calcul, et par $f(a, b)$ le résultat de ce calcul approché, la formule (8) donne le terme de correction

$$\left[h f'_a(a, b) + k f'_b(a, b)\right],$$

qui est de l'ordre des premières puissances de h et de k ; et, de plus, l'erreur que cette correction laisse encore subsister dans la valeur de $f(a + h, b + k)$, est exprimée par le

reste R qui est du second ordre, c'est-à-dire ne contient que des termes du second degré en h et k.

Si h et k sont très-petits, le rapport du reste R au terme du premier ordre

$$\left[h f'_a(a, b) + k f'_b(a, b)\right]$$

sera aussi très-petit, et il aura pour limite zéro en même temps que h et k. En effet, si l'on pose $k = \omega h$, ce rapport prend la forme

$$\frac{\frac{h}{2}\left[f''_{a,a}(a+\theta h, b+\theta k) + 2\omega f''_{a,b}(a+\theta h, b+\theta k) + \omega^2 f''_{b,b}(a+\theta h, b+\theta k)\right]}{f'_a(a, b) + \omega f'_b(a, b)};$$

ω est arbitraire, mais fini; et lorsque h est très-petit, on voit que le numérateur de la fraction ci-dessus est très-petit, et qu'il a zéro pour limite en même temps que h, tandis que le dénominateur reste fini.

Il suit de là que, pour de très-petites valeurs de h et k, l'accroissement de la fonction $f(a, b)$ ou

$$f(a+h, b+k) - f(a, b),$$

est sensiblement égal à $h f'_a(a, b) + k f'_b(a, b)$, c'est-à-dire composé de termes *proportionnels aux accroissements des variables*, et égaux respectivement au produit de chaque accroissement par la dérivée partielle de la fonction, relative à la variable qui a reçu cet accroissement.

Ce théorème est vrai pour un nombre quelconque de variables indépendantes. Il suppose toutefois que les dérivées partielles $f'_a(a, b)$, $f'_b(a, b)$ ne soient pas nulles à la fois; autrement le terme du premier ordre

$$h f'_a(a, b) + k f'_b(a, b)$$

disparaîtrait. Ce cas d'exception correspond à celui que

nous avons déjà signalé au n° 54 pour les fonctions d'une seule variable; et il se présente, comme ce dernier, toutes les fois que la fonction $f(x, y)$ passe pour les valeurs $x = a$, $y = b$, par un maximum ou un minimum.

72. Les formules relatives au calcul approché des fonctions

$$\frac{1}{1+x}, \quad \sqrt{1+\varepsilon}, \quad \sqrt[3]{1+\varepsilon}, \ldots, \quad \sqrt[n]{1+\varepsilon},$$

dont il a été question dans les nos 32, 44 et 45, sont évidemment des cas particuliers de l'équation (7). Bornons-nous à rechercher la limite supérieure de l'erreur que l'on fait en remplaçant

$$\sqrt[n]{1+\varepsilon} \text{ par } 1 + \frac{\varepsilon}{n},$$

limite dont nous avons ajourné l'expression (45). A cet effet, posons dans l'équation (7)

$$f(a) = \sqrt[n]{a},$$

d'où

$$f'(a) = \frac{1}{n\sqrt[n]{a^{n-1}}} \quad \text{et} \quad f''(x) = -\frac{n-1}{n^2\sqrt[n]{a^{2n-1}}};$$

puis $a = 1$, $h = \varepsilon$. Il viendra

$$\sqrt[n]{1+\varepsilon} - 1 - \frac{\varepsilon}{n} = -\frac{(n-1)\varepsilon^2}{2n^2\sqrt[n]{(1+\theta\varepsilon)^{2n-1}}}.$$

Réduisons $1 + \theta\varepsilon$ à l'unité dans le second membre, et nous aurons, pour limite supérieure de l'erreur,

$$\frac{n-1}{2n^2}\varepsilon^2.$$

Le signe du second membre nous apprend en outre que $1 + \frac{\varepsilon}{n}$ est approché par excès.

Pour $n = 2$, on retrouve $\frac{\varepsilon^2}{8}$, comme au n° 44.

Pour $n = 3$, on a $\frac{\varepsilon^2}{9}$, limite préférable à celle qu'avaient donnée les transformations effectuées au n° 45.

73. L'équation (7) fournit encore très-simplement des limites supérieures de l'erreur qui provient de la proportion tabulaire dans les calculs logarithmiques; et même on va voir que ces limites sont plus approchées que celles que nous avons tirées de l'équation (1).

1°. Soit

$$f(x) = \log x, \quad \text{d'où} \quad f'(x) = \frac{\log e}{x}, \quad f''(x) = -\frac{\log e}{x^2}.$$

Il vient

$$\log(a+h) - \log a \quad \text{ou} \quad \delta = \frac{h \log e}{a} - \frac{h^2 \log e}{2(a+\theta h)^2},$$

$$\log(a+1) - \log a \quad \text{ou} \quad \Delta = \frac{\log e}{a} - \frac{\log e}{2(a+\theta_1)^2},$$

et, par suite,

$$\delta - h\Delta = \frac{\log e}{2}\left[\frac{h}{(a+\theta_1)^2} - \frac{h^2}{(a+\theta h)^2}\right].$$

La valeur numérique du second membre est l'erreur commise en prenant $\delta = h\Delta$. Or θ, θ_1 étant positifs et h moindre que 1, le facteur entre parenthèses est, en valeur absolue, plus petit que $\frac{1}{a^2}$; d'ailleurs on a $\log e < \frac{1}{2}$; donc l'erreur de la proportion tabulaire est moindre que $\frac{1}{4a^2}$; et en particulier pour un nombre a supérieur à 10000, l'erreur est moindre que le quart d'une unité du huitième ordre décimal. Cette limite est moitié de celle que la formule (1) nous avait fournie (61), et il serait inutile d'en chercher une plus approchée, puisqu'on ne conserve que sept décimales dans les Tables.

2°. Soit

$$f(x) = \log \sin x, \quad \text{d'où} \quad f'(x) = \frac{\log e}{\operatorname{tang} x}, \quad f''(x) = -\frac{\log e}{\sin^2 x}.$$

Il vient, en continuant à désigner par δ et Δ les différences

$$\log \sin(a + h'') - \log \sin a \quad \text{et} \quad \log \sin(a + 10'') - \log \sin a,$$

$$\delta = \frac{h \log e}{\operatorname{tang} a} \cdot \operatorname{arc} 1'' - \frac{h^2 \log e}{2 \sin^2(a + \theta h)} \cdot \operatorname{arc}^2 1'',$$

$$\Delta = \frac{10 \log e}{\operatorname{tang} a} \cdot \operatorname{arc} 1'' - \frac{10^2 \log e}{2 \sin^2(a + \theta_1 10'')} \cdot \operatorname{arc}^2 1''.$$

(Pour plus de clarté, nous mettons en évidence, comme au n° 63, le facteur arc $1''$, en sorte que la lettre h, hors des signes trigonométriques, désigne un nombre abstrait.)

Des deux équations précédentes, on tire

$$\delta - \frac{h}{10}\Delta = \frac{\log e . \operatorname{arc}^2 1''}{2}\left[\frac{10 h}{\sin^2(a + \theta_1 10'')} - \frac{h^2}{\sin^2(a + \theta h)}\right].$$

θ, θ_1 étant positifs et h moindre que 10, le facteur entre parenthèses du second membre est, en valeur absolue, moindre que $\frac{100}{\sin^2 a}$. Mais, au lieu d'adopter cette limite analogue à celle du cas précédent, il sera bon ici de montrer qu'on peut la réduire de moitié.

En effet, si l'on pose

$$h = \frac{10}{p}, \quad \left[\frac{\sin a}{\sin(a + \theta_1 10'')}\right]^2 = \alpha, \quad \left[\frac{\sin a}{\sin(a + \theta h)}\right]^2 = \beta,$$

le facteur dont il s'agit prend la forme

$$\frac{100}{\sin^2 a} \cdot \frac{\alpha p - \beta}{p^2}.$$

α et β sont extrêmement peu inférieurs à 1; car l'un et l'autre surpassent $\left[\frac{\sin a}{\sin(a + 10'')}\right]$; et cette fraction, qui

croît avec a, prend pour $a = 5^\circ$, la valeur

$$\left[\frac{\sin 5^\circ}{\sin(5^\circ + 10'')}\right]^2 = 0,9988\ldots.$$

Ainsi α et β sont compris entre 1 et 0,998, pour tout arc égal ou supérieur à 5 degrés.

Maintenant, on s'assure aisément que l'inégalité

$$\frac{\alpha p - \beta}{p^2} < \frac{1}{2} \quad \text{(où l'on suppose } \alpha p > \beta\text{)}$$

est vraie; car elle équivaut à

$$(p - \alpha)^2 + (2\beta - \alpha^2) > 0,$$

dont le premier membre est essentiellement positif, puisque 2β surpasse 1, et par suite, α^2. Mais il faut aussi vérifier l'inégalité

$$\frac{\beta - \alpha p}{p^2} < \frac{1}{2} \quad \text{(où l'on suppose } \alpha p < \beta\text{)}.$$

Il suffit de changer les signes de α et β dans la précédente, ce qui donne

$$(p + \alpha)^2 - (2\beta + \alpha^2) > 0.$$

Or p étant plus grand que 1, on a

$$(p + \alpha)^2 > (1,998)^2 > 3 \quad \text{et} \quad 2\beta + \alpha^2 < 2 + 1 = 3.$$

Ainsi la valeur numérique du rapport $\frac{\alpha p - \beta}{p^2}$ est toujours moindre que $\frac{1}{2}$; par suite, celle du facteur

$$\frac{10h}{\sin^2(a + \theta_1 10'')} - \frac{h^2}{\sin^2(a + \theta h)}$$

est plus petite que $\frac{50}{\sin^2 a}$.

Donc enfin la différence $\delta - \frac{h}{10}\Delta$, c'est-à-dire l'erreur qui résulte de la proportion tabulaire est plus petite que $\frac{50 \operatorname{arc}^2 1''}{4 \sin^2 a}$ ou $\frac{\operatorname{arc}^2 10''}{8 \sin^2 a}$; et, en remplaçant arc $10''$ par $\frac{1}{2.10^4}$,

on a

$$\frac{1}{32\ 10^8.\sin^2 a}$$

pour limite supérieure de l'erreur que la proportion comporte. Elle est le quart de celle que nous avions tirée de la formule (1); et c'est avec elle qu'on voit bien pourquoi l'on peut faire usage avec confiance de la règle des parties proportionnelles, pour un arc égal ou supérieur à 5 degrés. En effet, nous avons remarqué au n° 63 que la fraction précédente se réduit alors à $\frac{1}{10^7.2,4}$ environ, et, par suite, à partir de l'arc de 5 degrés, l'erreur est moindre que la moitié d'une unité du septième ordre décimal.

Dans les applications qui précèdent, la formule (7) a une supériorité marquée, pour l'évaluation de l'erreur, sur la formule (1). On en verra de nouveaux exemples dans ce qui va suivre. Cependant, comme la formule (1) n'exige que la connaissance de la dérivée première $f'(x)$, et qu'elle suffit dans un grand nombre de cas, il était utile d'en connaître l'usage pour apprécier le degré d'approximation, *sans recourir à la dérivée seconde;* c'est pourquoi nous avons montré avec détails, dans les n°s 61, 62, 63 et 64, le parti qu'on peut tirer de cette formule, et nous suivrons la même marche dans le calcul approché des racines des équations numériques qui terminera cette Théorie.

Le lecteur sera ainsi à même de comparer les deux procédés.

74. Comme application de la formule (8), où l'on considère deux variables indépendantes, proposons-nous de rechercher quel est, dans la formation d'un réseau géodésique, le triangle *le plus avantageux,* c'est-à-dire celui pour lequel les erreurs inévitables commises dans l'observation des angles influent le moins possible sur les deux côtés à calculer.

Soit b le côté servant de base, et que l'on suppose mesuré ou calculé avec assez de soin pour que l'erreur soit négligeable; A, B, C les angles dont les mesures comportent de petites erreurs h, k, l, en sorte que les angles observés sont $A+h$, $B+k$, $C+l$; enfin, a et b les deux côtés cherchés. On a

$$a=\frac{b\sin A}{\sin B},\quad c=\frac{b\sin C}{\sin A}.$$

Mais la valeur du côté a, telle qu'elle résulte de l'observation, est

$$\frac{b\sin(A+h)}{\sin(B+k)}.$$

L'erreur absolue commise sur ce côté, est donc

$$\varepsilon=b\left[\frac{\sin(A+h)}{\sin(B+k)}-\frac{\sin A}{\sin B}\right].$$

Comme h et k sont très-petits par rapport à A et B, le théorème du n° 71 est applicable, et l'on a sensiblement

$$\varepsilon=b\left[h\frac{\cos A}{\sin B}-k\frac{\sin A\cos B}{\sin^2 B}\right];$$

par conséquent, l'erreur relative est

$$\frac{\varepsilon}{a}=h\cot A-k\cot B.$$

Il est permis de regarder les erreurs h et k (qui sont indépendantes des grandeurs des angles A et B) comme sensiblement égales et de même sens. On aura donc simplement, en faisant $h=k$,

$$\frac{\varepsilon}{a}=h(\cot A-\cot B);$$

d'où l'on voit que l'*erreur s'évanouit si l'on suppose* $A=B$. Ainsi, dans ce cas, bien que A et B soient fautifs, la valeur calculée du côté a sera exacte. Si l'égalité *mathématique* de h et k n'a pas lieu, le rapport $\frac{\varepsilon}{a}$ ne sera pas rigou-

reusement nul; mais du moins il sera très-voisin de zéro, pour $A = B$. Ce que nous venons de dire du côté a s'applique au côté c. Il résulte de là que le triangle le plus avantageux est celui dans lequel on a

$$A = B = C,$$

c'est-à-dire le triangle équilatéral.

Calcul approché des racines des équations numériques.

75. Nous terminerons notre théorie des approximations numériques en appliquant au calcul approché des racines des équations les formules (1) et (7) des nos 50 et 67.

Soit $f(x) = 0$ une équation quelconque qui n'a pas de racines égales. Comme en changeant x en $-x$ on ramène la recherche des racines négatives à celle des racines positives, nous supposerons qu'il s'agisse de calculer une racine positive de cette équation. Soit a une valeur approchée déjà connue, et h le nombre (positif ou négatif) qu'il faut ajouter à a pour avoir la racine exacte, on aura

$$f(a + h) = 0.$$

En admettant que a soit assez approché de la racine pour que la dérivée $f'(x)$ reste finie et continue dans le petit intervalle qui sépare a de $a + h$, la formule (1) sera applicable et donnera

$$(9) \qquad h = \frac{-f(a)}{f'(a + \theta h)}.$$

Supposons que a soit approché *par défaut*; h sera positif, et, par suite, $f(a)$ et $f'(a + \theta h)$ seront de signes contraires. Soit, pour fixer les idées, $f(a) < 0$, d'où $f'(a + \theta h) > 0$.

Pour tirer de l'équation (9) une valeur approchée de h,

on est conduit à remplacer $f'(a+\theta h)$, dont la valeur n'est pas connue, par un nombre en excès ou en défaut. Ce nombre se présente de lui-même dans les applications; parce que, avant de procéder au calcul d'une racine par approximations successives, on a dû débuter par la *séparer*, c'est-à-dire par déterminer *deux* nombres qui comprennent cette racine et la comprennent seule. L'un de ces nombres est celui que nous avons désigné par a. Soit b l'autre, qui surpasse la racine. $f(a)$ et $f(b)$ sont de signes contraires.

On peut admettre que b est assez rapproché de a pour que $f'(x)$ soit constamment croissante (ou décroissante) entre $x=a$ et $x=b$; de plus, comme l'équation proposée n'a pas de racines égales, la racine cherchée n'annule pas $f'(x)$, et par suite cette dérivée ne change pas de signe entre a et b, si ces deux nombres sont suffisamment voisins l'un de l'autre.

Cela posé, *si* $f'(x)$ *est croissante*, on aura

$$f'(a)<f'(a+\theta h)<f'(b),$$

d'où

$$\frac{-f(a)}{f'(a)}>h>\frac{-f(a)}{f'(b)}.$$

Si $f'(x)$ *est décroissante*, les signes d'inégalité seront renversés. Dans les deux cas, on peut dire que les décimales communes aux deux fractions

$$\frac{-f(a)}{f'(a)},\quad \frac{-f(a)}{f'(b)}$$

appartiendront à la vraie valeur de h.

76. Il est aisé d'interpréter ces résultats par les considérations géométriques auxquelles nous avons déjà eu recours dans le n° 51. Posons

$$y=f(x),$$

et regardons y et x comme les coordonnées rectangulaires

d'une courbe PmRn (*fig.* 3) rapportée aux deux axes

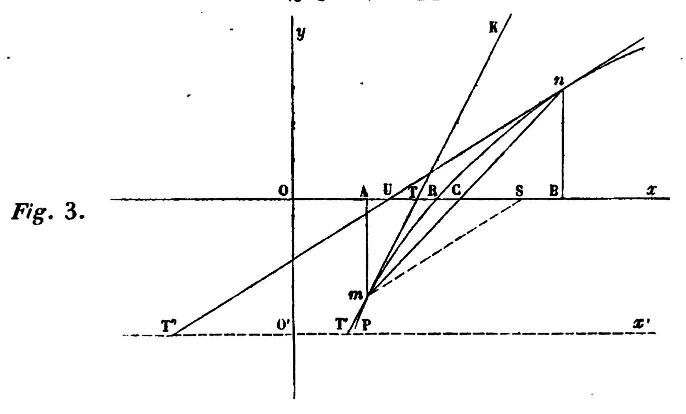


Fig. 3.

Ox, Oy. L'abscisse OR du point où la courbe va couper l'axe des x, représente la racine de l'équation $f(x) = 0$, qu'il s'agit de calculer. Soient

$$OA = a, \quad OB = b; \quad Am = -f(a), \quad Bn = f(b).$$

Si l'on mène la tangente au point m, laquelle rencontre en T l'axe des x, le triangle rectangle ATm donne

$$AT = \frac{mA}{\tang mTA} = \frac{mA}{\tang kTx} = \frac{-f(a)}{f'(a)}.$$

Puis, si l'on mène du même point m, une parallèle mS à la tangente nU qui répond au point n, le triangle rectangle mAS, donne

$$AS = \frac{mA}{\tang mSA} = \frac{mA}{\tang nUx} = \frac{-f(a)}{f'(b)}.$$

Ainsi AT et AS représentent les deux limites trouvées ci-dessus pour h. Or le point R, où la courbe coupe l'axe des x, tombe *entre* les points T et S, et, par conséquent, l'une de ces limites, AT, est par défaut, tandis que l'autre, AS, est par excès. Toutefois ceci demande quelques développements, attendu que la courbe peut prendre diverses formes entre les points m et n.

Et d'abord, nous avons admis que $f'(x)$ était constamment croissante (ou décroissante) entre a et b. Quel est le sens géométrique de cette supposition ? Pour le bien définir, il faut préciser ce qu'on entend par *concavité* ou *convexité* d'une courbe vers une droite.

77. Une courbe PmQ (*fig.* 4) est dite *concave* en un de ses

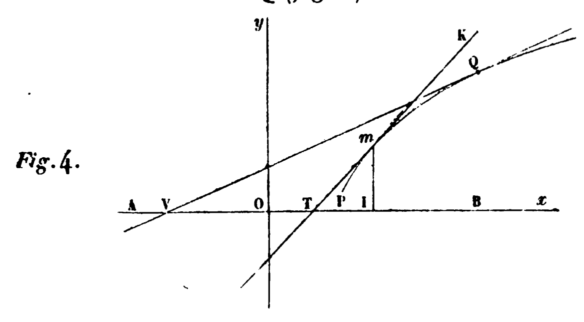


Fig. 4.

points m vers une droite AB, lorsque deux arcs très-petits de cette courbe (mP, mQ), comptés *de part et d'autre du point m*, sont situés *dans l'angle aigu mTB* que la tangente en m fait avec cette droite. Si, au contraire, ces deux arcs sont *tous deux dans l'angle obtus mTA* (*fig.* 5), la courbe

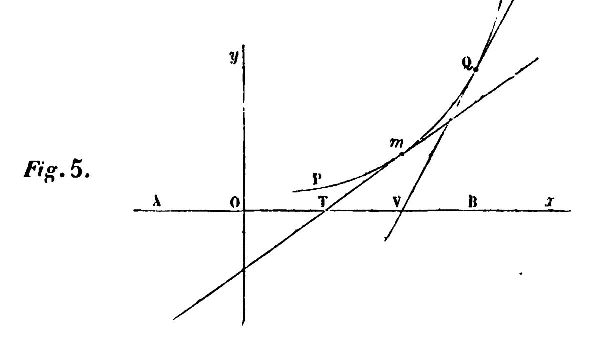


Fig. 5.

est dite *convexe* en m vers la droite AB.

Lorsqu'en suivant le cours continu d'une courbe PmQ, on rencontre un point m (*fig.* 6) tel, que la courbe, de

concave qu'elle était vers une droite AB dans la région Pm,

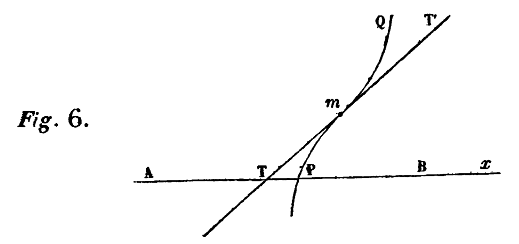


Fig. 6.

devient convexe dès qu'on atteint la région mQ, on dit qu'il y a *inflexion* en ce point. Un point d'inflexion est donc tel que les deux arcs de courbe (mP, mQ), comptés à partir de ce point, sont situés de côtés différents de la tangente TmT'; celle-ci traverse la courbe en même temps qu'elle la touche. Cela posé, reprenons la *fig.* 4, qui représente un arc PmQ, concave vers la droite AB, et situé tout entier d'un même côté de cette droite; prenons-la pour axe des x; plaçons l'origine des coordonnées rectangulaires en un point quelconque O de AB, et comptons les x positifs dans le sens OB et les y positifs du côté de l'arc PmQ. Soit $y = f(x)$ l'équation de cette courbe. Si nous faisons croître x à partir de l'abscisse OI qui répond au point m, ce point va marcher de m vers Q sur la courbe *au-dessous* de la droite mK prolongement de mT, et par suite il est visible que l'angle que fait la tangente avec l'axe des x va *décroître* depuis mTx jusqu'à QVx. Donc sa tangente trigonométrique, qui n'est autre que $f'(x)$, *est une fonction décroissante*.

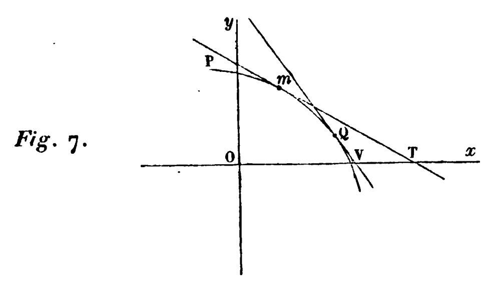


Fig. 7.

Si l'arc de courbe, toujours concave vers Ox, avait la dis-

position indiquée dans la *fig.* 7, où l'angle mTx est obtus, cet angle irait encore en décroissant de mTx à QVx, et sa tangente trigonométrique augmenterait en grandeur absolue; mais elle ne cesserait pas d'être une fonction décroissante, eu égard à son signe.

Concluons donc que *si une courbe dont l'ordonnée positive est* $y = f(x)$, *est concave en un de ses points vers l'axe des* x, *la dérivée* $f'(x)$ *sera décroissante à partir de la valeur de* x *qui répond à ce point.*

On démontrera de même, à l'aide de la *fig.* 5, que *si une courbe est convexe vers l'axe des* x, $f'(x)$ *sera croissante à partir de l'abscisse du point que l'on considère* (et dont l'ordonnée est toujours supposée positive).

Ces deux propositions directes ont leurs réciproques vraies : ce qu'on met en évidence par le raisonnement connu.

En résumé, le caractère analytique qui distingue un arc *concave* d'un arc *convexe* vers une droite prise pour axe des x (arc qu'on suppose situé tout entier dans la région des y positifs), c'est que la dérivée de l'ordonnée y prise par rapport à x, ou $f'(x)$, *est une fonction décroissante* dans le premier cas, et *croissante* dans le second. On s'assure d'ailleurs (48) que $f'(x)$ est continuellement croissante (ou décroissante) depuis une valeur $x = a$ jusqu'à $x = b$, en examinant si la dérivée seconde $f''(x)$ garde constamment le signe $+$ (ou le signe $-$) dans cet intervalle.

Enfin, le caractère analytique d'un *point d'inflexion* (*fig.* 6), c'est qu'en ce point la tangente trigonométrique de l'angle mTx ou $f'(x)$, passe par un maximum ou par un minimum; et par suite, lorsque x atteint et dépasse

l'abscisse d'un point d'inflexion, $f''(x)$ *doit changer de signe* (*).

78. Maintenant, rien n'est plus aisé que d'interpréter géométriquement l'hypothèse faite au n° 75, savoir que $f'(x)$ est constamment croissante (ou décroissante) entre $x = a$ et $x = b$. En effet, il suffit d'imaginer (*fig.* 3 du n° 76) une parallèle $O'x'$ à l'axe des x, laquelle laisse l'arc mn *au-dessus* d'elle. Alors l'hypothèse faite signifie que la courbe est de m en n, toujours convexe (ou toujours concave) vers cette parallèle inférieure, sans offrir d'inflexion.

Dans la *fig.* 3, c'est la *concavité* qui a lieu, mais il y a trois autres dispositions à considérer : elles sont représentées dans les *fig.* 8, 9 et 10.

Dans la *fig.* 8, il y a encore *concavité* vers une parallèle inférieure à l'axe des x, qui laisserait l'arc mn au-dessus d'elle. Mais ce qui distingue cette figure de la *fig.* 3, c'est que l'ordonnée $f(a)$ qui répond au point m est positive.

Dans les *fig.* 9 et 10 la courbe est *convexe* vers une parallèle inférieure à l'axe des x; et ce qui les distingue l'une de l'autre, c'est que l'ordonnée $f(a)$ du point m est négative dans la *fig.* 9 et positive dans la *fig.* 10.

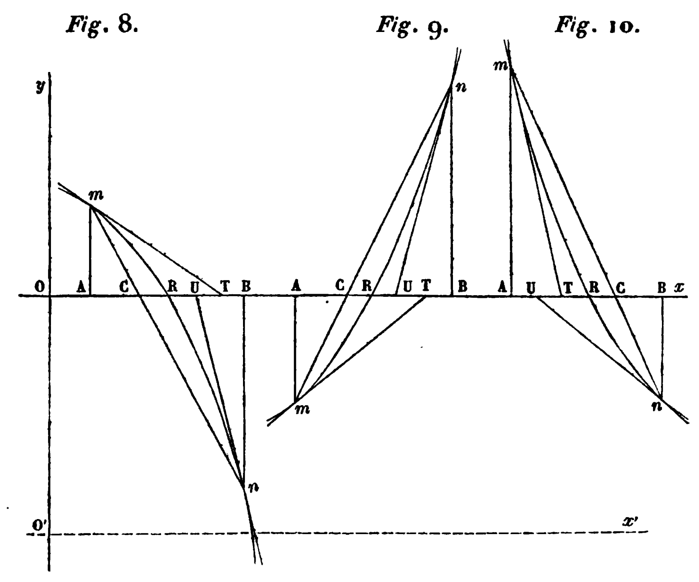
Il est clair qu'il n'y a pas d'autres combinaisons que les quatre que nous venons de figurer.

Or il est facile de s'assurer que les expressions trouvées pour AT et AS sur la *fig.* 3 conviennent, sans modifica-

---

(*) Les caractères analytiques de la concavité, de la convexité et de l'inflexion demeurent les mêmes quand la courbe dont l'équation est $y = f(x)$, est supposée rapportée à des axes de coordonnées *obliques*. La dérivée $f'(x)$ représente alors un rapport de sinus qui est croissant ou décroissant dans les mêmes circonstances que ci-dessus.

Fig. 8. *Fig.* 9. *Fig.* 10.



tion, aux trois autres cas. Ainsi, dans la *fig.* 8 par exemple, on a

$$AT = \frac{m\,A}{\text{tang}\,m\,TA} = \frac{-m\,A}{\text{tang}\,m\,Tx} = \frac{-f(a)}{f'(a)};$$

etc. Cette sous-tangente AT, ou $\frac{-f(a)}{f'(a)}$, est précisément le terme de correction qu'adopte Newton dans sa méthode d'approximation, et qu'il ajoute à a, en sorte qu'il prend pour valeur approchée de la racine $a - \frac{f(a)}{f'(a)}$.

Cette valeur est représentée par $OA + AT = OT$, dans les quatre figures considérées. Or il suffit d'y jeter les yeux pour reconnaître que le point T tombe tantôt en deçà, tantôt au delà du *point racine* R. Il est en deçà dans les *fig.* 3 et 10, et alors l'expression

$$(10) \qquad - \frac{f(a)}{f'(a)}$$

fournit une approximation par défaut, c'est-à-dire de même sens que la valeur a dont on est parti.

Il est au delà dans les *fig.* 8 et 9, et alors l'expression (10) fournit une approximation par excès.

Remarquons maintenant que les *fig.* 3 et 10 ont cela de commun, que $f(a)$ et $f''(a)$ y sont de même signe. En effet, dans la *fig.* 3 il y a concavité, et par suite, $f''(a) < 0$; d'ailleurs on a aussi $f(a) < 0$; donc $f(a)$ et $f''(a)$ sont de même signe. Dans la *fig.* 10 il y a convexité, et par suite $f''(a) > 0$; mais on a aussi $f(a) > 0$; donc $f(a)$ et $f''(a)$ sont encore de même signe.

Au contraire, dans chacune des *fig.* 8 et 9, $f(a)$ et $f''(a)$ sont de signes contraires.

Concluons donc que le *terme de correction de Newton*,

$$-\frac{f(a)}{f'(a)},$$

fournira une approximation par défaut de la racine, toutes les fois que f (a) *et* f'' (a) *seront de même signe, et une approximation par excès dans le cas contraire.*

Quant à l'autre terme

$$-\frac{f(a)}{f'(b)},$$

auquel nous sommes également parvenus dant le n° 75, et qui est représenté dans la *fig.* 3 par la longueur AS, on n'en fait aucun usage, bien qu'il fournisse visiblement une approximation de *sens contraire* à la précédente : nous allons donner le motif de ce rejet.

79. Reprenons l'une des *fig.* 3, 8, 9, 10; la *fig.* 3 par exemple, et tirons la sécante *mn* qui va couper l'axe des x au point C. Puisque l'arc *mn* est sans inflexion, cette corde laissera l'arc *tout entier* d'un même côté d'elle, et d'ailleurs elle passera évidemment *entre* les deux parallèles indéfinies mS, nT''. Par conséquent, le point C tombera toujours du *même côté de* R *que le point* S, et *entre* R *et* S. Il en résulte que si l'on calcule la sous-sécante AC, on aura une approximation toujours plus grande qu'avec AS, et

de même sens qu'elle. AC remplacera donc AS avec avantage. Or les triangles semblables donnent

$$\frac{AC}{AB}=\frac{Am}{Am+nB}, \quad \text{ou} \quad \frac{AC}{b-a}=\frac{-f(a)}{-f(a)+f(b)},$$

d'où

$$(11) \qquad AC=-\frac{(b-a)f(a)}{f(b)-f(a)}.$$

Les mêmes triangles semblables existent dans les *fig.* 8, 9 et 10, et conduisent à la même expression de la sous-sécante AC, en ayant égard aux signes des ordonnées extrêmes.

Il nous reste une remarque importante à faire : c'est que la formule (11) n'est autre que celle qu'on avait déjà trouvée au n° 56, en supposant les accroissements du premier membre de l'équation proportionnels aux accroissements de l'inconnue. Il n'y a que les notations de changées : ce que nous appelions A est ici désigné par $f(a)$; h est $b-a$, et ΔA est $f(b)-f(a)$. Ainsi *la règle de fausse position* ou *des parties proportionnelles* revient à considérer la courbe $y=f(x)$ comme confondue, dans l'intervalle des valeurs a et b attribuées à x, avec la ligne droite qui joint les deux points correspondants à ces valeurs.

En définitive, on adoptera pour le calcul approché de la racine comprise entre a et b, l'un ou l'autre des deux termes

$$(10) \qquad -\frac{f(a)}{f'(a)}, \qquad (11) \qquad -\frac{(b-a)f(a)}{f(b)-f(a)};$$

et si on les calcule tous deux, comme l'un est par défaut tandis que l'autre est par excès, on sera sûr que l'erreur commise sera inférieure à la valeur absolue de leur différence.

Ces considérations suffisent déjà pour rectifier complétement la méthode d'approximation de Newton, en faisant connaître, pour chaque valeur approchée, *le sens de l'erreur commise et une limite supérieure de cette erreur.*

Si l'on prenait la moyenne arithmétique entre les deux termes (10) et (11) on aurait une limite d'erreur moindre encore, et marquée par leur demi-différence; mais on ne saurait plus si le résultat obtenu en ajoutant cette moyenne à a, serait approché par défaut ou par excès.

80. *Application.* — Soit l'équation

$$x^3 - x^2 - 2x + 1 = 0,$$

dont le premier membre sera désigné par $f(x)$. Supposons que, par des substitutions successives de nombres croissants de dixièmes en dixièmes, on ait reconnu que cette équation a une racine positive comprise entre 1,8 et 1,9.

En posant

$$a = 1,8, \quad b = 1,9,$$

on aura

$$f(a) = -0,008, \quad f(b) = +0,449.$$

La dérivée

$$f'(x) = 3x^2 - 2x - 2$$

ne change pas de signe pour les valeurs de x comprises entre 1,8 et 1,9. De plus, $f''(x)$ ou $6x - 2$ est évidemment positive dans le même intervalle. Ainsi $f(a)$ et $f''(a)$ sont de signes contraires. C'est le cas de la *fig.* 9.

Le terme de correction de Newton $-\dfrac{f(a)}{f'(a)}$ sera donc par excès (78); et par suite, le terme fourni par la règle des parties proportionnelles $-\dfrac{(b-a)f(a)}{f(b)-f(a)}$ sera par défaut. Ainsi, en développant le premier terme en décimales, on devra prendre le quotient par excès; et si l'on a recours au second, on devra le prendre par défaut.

Il vient

$$\frac{-f(a)}{f'(a)} = \frac{0,008}{4,12} = 0,00195 \text{ (par excès)},$$

$$\frac{-(b-a)f(a)}{f(b)-f(a)} = \frac{0,1 \times 0,008}{0,457} = 0,00175 \text{ (par défaut)}.$$

Ajoutant successivement ces quotients à $a = 1,8$, on aura deux nouvelles valeurs a_1, b_1,

$$a_1 = 1,80175, \quad b_1 = 1,80195,$$

entre lesquelles la racine sera comprise. Elle sera donc déjà connue avec trois décimales exactes, et l'on saura que l'erreur commise, en prenant a_1 ou b_1, est moindre que 2 dix-millièmes : on verra plus loin que l'approxima ion de b_1 est encore plus grande que ne l'indique cette limite.

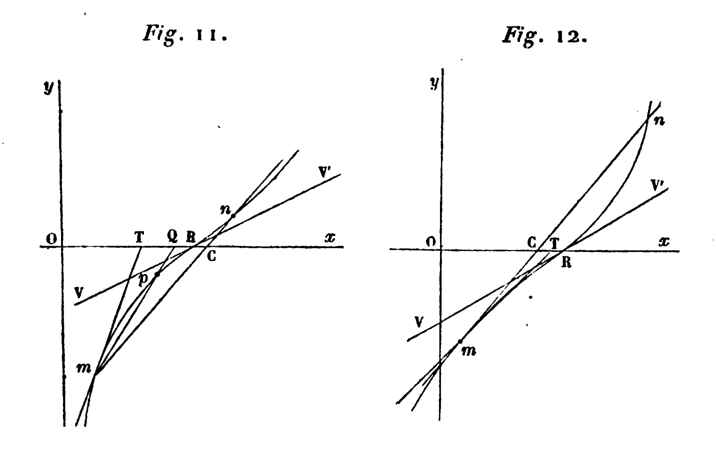
81. On s'assure aisément si $f''(x)$ garde constamment le même signe entre a et b, toutes les fois que cette dérivée seconde peut se mettre sous la forme $\varphi(x) - \psi(x)$, $\varphi(x)$ et $\psi(x)$ étant des fonctions positives et croissantes. [C'est ce qui a toujours lieu, par exemple, dans le cas d'une *fonction algébrique*, $\varphi(x)$ désignant le groupe des termes positifs, et $\psi(x)$ celui des termes négatifs.] En effet, puisque l'on a $f''(x) = \varphi(x) - \psi(x)$, la valeur que prend $f''(x)$, quand x varie entre a et b, est toujours comprise entre les deux différences

$$\varphi(a) - \psi(b) \quad \text{et} \quad \varphi(b) - \psi(a);$$

donc, si elles sont de même signe, on est certain que $f''(x)$ garde aussi le même signe entre a et b. Cette condition est suffisante, mais elle n'est pas nécessaire.

Nota. — Nous avons dit que l'on pouvait toujours admettre la permanence du signe de $f''(x)$ entre a et b, pourvu que b fût suffisamment voisin de a. Il y a cependant un cas d'exception : c'est celui où $f''(x)$ passerait précisément par zéro, en changeant de signe, pour la valeur de x égale à la racine cherchée. Alors la courbe $y = f(x)$ présenterait une inflexion au point R, comme le montrent

les *fig.* 11 et 12. La tangente VV′ au point R traverse la



courbe, en laissant une portion mR au-dessous d'elle et l'autre Rn au-dessus ; en sorte que l'on ne prévoit plus si le point C où la corde mn va couper l'axe des x, est en deçà ou au delà du point R. Ainsi, dans la *fig.* 11, il est au delà, et T et C tombent encore de côtés différents de R comme dans le cas général. Au contraire, dans la *fig.* 12, C et T sont d'un même côté de R.

Mais on peut remarquer que l'incertitude sur la position du point C n'existerait plus si, au lieu de partir de deux valeurs de x, a et b, *entre lesquelles* tombe la racine cherchée, on partait, dans ce cas, de valeurs *toutes deux par défaut* (ou *par excès*), par exemple celles qui répondent à deux points m et p de la courbe, situés en deçà de R. Alors la sécante mp couperait l'axe des x en un point Q *toujours en deçà* de R (ou *toujours au delà* de R, si l'arc mR était convexe) ; avec cette précaution, la règle des parties proportionnelles donnerait donc une approximation dont le sens serait encore prévu ; seulement ce sens ne serait plus opposé à celui de l'approximation fournie par la méthode de

Newton, et par suite les deux méthodes ne pourraient plus être combinées comme au nº 79. Nous donnerons un exemple de ce cas singulier dans la résolution d'une équation transcendante (96).

Si l'équation $f(x) = 0$ est algébrique, on sortira aisément de la difficulté, en remarquant qu'il doit alors exister un plus grand commun diviseur entre $f(x)$ et $f''(x)$, puisque ces deux polynômes sont annulés par la même valeur de x. La recherche de la racine se fera donc sur l'équation de degré moindre qu'on obtient en égalant à zéro ce plus grand commun diviseur.

82. Nous avons supposé jusqu'ici, pour fixer les idées, que l'on partait d'une valeur a approchée *par défaut*, mais la formule (9) du nº 75 n'implique point cette restriction. On peut y remplacer a par la valeur en excès b, et il viendra

$$h = -\frac{f(b)}{f'(b+\theta h)},$$

seulement cette valeur de h sera maintenant négative. En négligeant θh au dénominateur, on aura une valeur approchée de h,

$$-\frac{f(b)}{f'(b)};$$

c'est le *terme de correction de Newton*, appliqué à la valeur par excès b. Ce terme est évidemment représenté en valeur absolue par la sous-tangente BU (*fig.* 3, 8, 9 et 10). Cette sous-tangente retranchée de OB fournira une longueur

$$\text{OU} = b - \frac{f(b)}{f'(b)},$$

qui approchera tantôt par défaut, tantôt par excès de OR; par défaut dans les *fig.* 3 et 10, par excès dans les deux autres.

En même temps, si l'on change a en b, et réciproquement, dans la formule (11), la sous-sécante AC se changera dans BC, et l'on aura

$$BC = -\frac{(b-a)f(b)}{f(b)-f(a)}.$$

C'est le terme de correction que fournit la règle des parties proportionnelles appliquée à la valeur par excès b.

Il est visible que les points U et C tombent toujours de côtés différents du *point racine* R. Ainsi tout ce que nous avons dit de la valeur par défaut a s'applique à la valeur par excès b.

83. Maintenant il est aisé de répondre à cette question : *De laquelle des valeurs initiales* a *ou* b *faut-il partir pour que, substituée à la place de* x *dans la formule*

$$-\frac{f(x)}{f'(x)},$$

elle fournisse une approximation constamment de même sens que cette valeur, et, par suite, *constamment croissante?*

L'inspection des quatre *fig.* 3, 8, 9, 10 montre que le point T est du même côté de R que le point A, *seulement* dans les *fig.* 3 et 10, où $f(a)$ et $f''(a)$ sont de même signe; et le point U est du même côté de R que le point B, *seulement* dans les *fig.* 8 et 9, où $f(b)$ et $f''(b)$ sont aussi de même signe.

Concluons donc que *la valeur* a *ou* b *dont on doit partir pour que l'approximation de Newton soit toujours croissante et de même sens que cette valeur, est celle qui, mise à la place de* x *dans* f(x) *et* f''(x), *donne des résultats de même signe.*

84. Puisqu'on a vu plus haut que *la règle des parties proportionnelles* donnait, dans les mêmes circonstances, une approximation toujours de *sens contraire à celle de*

Newton, il est clair que si l'on veut employer *exclusivement* au calcul approché d'une racine la règle des parties proportionnelles, il faudra renverser la conclusion précédente, et dire :

La valeur *a* ou *b* *dont on doit partir pour que l'approximation fournie par la règle des parties proportionnelles soit toujours croissante et de même sens que cette valeur, est celle qui, mise à la place de* x *dans* f (x) *et* f'' (x), *donne des résultats de signes contraires.*

Dans les équations *algébriques*, les calculs de substitution qu'exige la méthode de Newton sur la dérivée $f'(x)$, sont ordinairement assez simples; c'est pourquoi on fait un grand usage de cette méthode. Mais, dans la résolution des équations *transcendantes*, où entrent par exemple des logarithmes de lignes trigonométriques, il peut arriver que les calculs à effectuer sur la dérivée $f'(x)$ se compliquent; alors on préférera employer la règle des parties proportionnelles, où les calculs ne portent que sur la fonction primitive. Nous donnerons des exemples des deux méthodes sur quelques équations transcendantes.

85. *Remarque.* — En formulant les conclusions précédentes, nous n'avons pas voulu dire que, dans la méthode de Newton ou des parties proportionnelles, on ne devait *jamais* partir de celle des deux valeurs *a* ou *b* pour laquelle l'approximation change de sens. L'essentiel est que ce sens soit nettement défini, afin qu'on ne marche pas au hasard. Si, par exemple, on a lieu de croire que *a* est beaucoup plus approchée de la racine que *b*, il pourra être plus avantageux d'appliquer la méthode à *a*, lors même que le sens de l'approximation serait renversé. On en voit un exemple dans l'équation traitée au n° 80. Si l'on partait de la valeur $b = 1,9$, pour laquelle $f(b)$ et $f''(b)$ sont de même signe, la formule de Newton donnerait la valeur par

excès

$$1,9 - \frac{0,449}{5,03} = 1,811\ldots.$$

Or elle est beaucoup moins approchée que la valeur également par excès

$$b_1 = 1,80195,$$

déjà trouvée, en partant de $a = 1,8$.

Si l'on veut approcher davantage de la racine, on partira maintenant de la valeur $b_1 = 1,80195$, et le terme de correction

$$-\frac{f(b_1)}{f'(b_1)}$$

fournira une nouvelle approximation de même sens que b_1 (83). On trouve

$$\frac{f(b_1)}{f'(b_1)} = \frac{0,0000507384\ldots}{4,13718\ldots} = 0,00001226, \text{ (quotient par défaut)}.$$

Et, par suite, on a pour seconde approximation par excès,

$$b_2 = b_1 - \frac{f(b_1)}{f'(b_1)} = 1,80193774.$$

On voit que cette correction n'a pas atteint les quatre premières décimales de b_1, et c'est ce qui nous a autorisés (comme on le verra plus bas) à pousser le quotient jusqu'à la huitième décimale. Cependant nous ne sommes pas encore sûrs de l'exactitude de tous ces chiffres.

Pour les vérifier, ce qu'il y a de plus court à faire, c'est la substitution immédiate dans l'équation proposée du nombre 1,80193773, qui n'est inférieur à b_2 que d'une unité du dernier ordre. Si le résultat de la substitution a le signe —, on sera certain que les huit chiffres trouvés sont exacts, attendu que l'on sait d'avance que la substitution de b_2 donnerait le signe +. C'est ce qui a lieu en effet. En substituant 1,80193773 on trouve — 0,00000002.

La règle des parties proportionnelles, si nous la faisions marcher de pair avec la méthode de Newton, nous fournirait

aussi un critérium certain. Car le terme de correction

$$-\frac{(b_1-a_1)f(b_1)}{f(b_1)-f(a_1)},$$

dont la valeur absolue est représentée dans la *fig.* 9 par la sous-sécante BC donne, comme on l'a déjà vu (79), une approximation de sens contraire à celle de la sous-tangente BT. Par conséquent les décimales communes aux deux résultats

$$b_1-\frac{f(b_1)}{f'(b_1)},\quad b_1-\frac{(b_1-a_1)f(b_1)}{f(b_1)-f(a_1)},$$

conviendront à la vraie valeur de x. Nous ne nous arrêterons pas à ce moyen de vérification.

86. Laissant de côté toutes considérations géométriques, nous allons rechercher, sur la formule (7), l'expression générale en grandeur et en signe de l'erreur que l'on commet en employant le terme de correction de Newton

$$-\frac{f(a)}{f'(a)}.$$

Nous ne ferons d'abord aucune hypothèse sur le sens de l'approximation de a.

$a+h$ désignant la valeur exacte d'une racine de l'équation $f(x)=0$, on aura

$$f(a+h)=0;$$

et, d'après la formule (7),

$$(12)\qquad h=-\frac{f(a)}{f'(a)}-\frac{h^2}{1.2}\cdot\frac{f''(a+\theta h)}{f'(a)}.$$

Donc l'erreur commise en prenant, pour seconde approximation de la racine,

$$a-\frac{f(a)}{f'(a)},$$

sera la valeur absolue de l'expression

$$\frac{h^2}{1.2}\cdot\frac{f''(a+\theta h)}{f'(a)}.$$

Si l'on suppose maintenant que a et b soient deux nom-

bres, le premier inférieur et le second supérieur à la racine $a+h$, et assez rapprochés pour que $f''(x)$ ne change pas de signe entre a et b, l'équation (12) montre immédiatement que h surpassera $\frac{-f(a)}{f'(a)}$, pourvu que $f'(a)$ et $f''(a)$ soient de signes contraires; ou, ce qui revient au même, pourvu que $f(a)$ et $f''(a)$ soient de même signe (*). Dans ce cas, $a-\frac{f(a)}{f'(a)}$ sera une valeur approchée dans le même sens que a. En remplaçant dans l'équation (12) a par b et h par $-k$, on voit de même que k surpassera $\frac{f(b)}{f'(b)}$, pourvu que $f(b)$ et $f''(b)$ soient de même signe; et, dans ce cas, $b-\frac{f(b)}{f'(b)}$ sera une valeur approchée dans le même sens que b. Ces conclusions avaient déjà été établies dans le n° 83 par la Géométrie.

87. Désignons par M un nombre égal ou supérieur à la plus grande des valeurs numériques que prend $f''(x)$ entre a et b, et par N un nombre égal ou inférieur à la plus petite valeur numérique de $f'(x)$ dans le même intervalle (nous avons donné précédemment des moyens de calculer ces deux nombres);

$$\frac{M}{2N}h^2$$

sera une limite supérieure de l'erreur commise. La fraction $\frac{M}{2N}$ sera calculée une fois pour toutes, et servira de coefficient à h^2 dans tout le cours du calcul; comme h n'est pas connu exactement, on le remplacera par un nombre supé-

(*) On sait, en effet, que $f(a)$ et $f'(a)$ sont de signes contraires. Rappelons en deux mots la raison de cette opposition de signe. Si $f'(a)$ est positive, la fonction $f(x)$ est croissante à partir de $x=a$; or elle s'annule pour $x=a+h$; donc elle doit être négative pour $x=a$. On démontre de même que si $f'(a)$ est négative, $f(a)$ sera positive.

rieur. Au début du calcul, ce nombre pourra être $b-a$, à défaut d'un autre plus approché.

Si l'on a $\frac{M}{2N}<1$, et que l'on parte, comme à l'ordinaire, d'une valeur a approchée à moins de $\frac{1}{10}$, le nombre des décimales exactes ira au moins en doublant à chaque approximation. Ceci peut encore avoir lieu, *lors même que* $\frac{M}{2N}$ *serait* >1, à cause de la petitesse du facteur inconnu h^2. Mais il ne faut pas oublier qu'à l'erreur théorique dont nous venons d'assigner une limite supérieure, s'ajoutent, dans la pratique, des erreurs provenant des décimales négligées dans le calcul des termes de correction.

Appliquons ces remarques à l'exemple du n° 80. En partant de la valeur par défaut $a=1,8$, nous avons trouvé pour h la valeur en excès

$$\frac{-f(1,8)}{f'(1,8)}=\frac{0,008}{4,12}<0,002.$$

Ce nombre 0,002 pourra donc être pris pour h dans l'évaluation de la limite supérieure de l'erreur. D'ailleurs on a

$$\frac{f''(a+\theta h)}{f'(a)}<\frac{f''(1,9)}{f'(1,8)}=\frac{9,4}{4,12};$$

ainsi le rapport que nous désignions plus haut par $\frac{M}{2N}$ est ici $\frac{9,4}{2\times 4,12}$; on n'a donc pas $\frac{M}{2N}<1$. Néanmoins on peut reconnaître, eu égard à la petitesse de h dont la valeur *par excès* est 0,002, que le nombre des décimales exactes a plus que doublé dans cette première approximation. En effet, l'erreur commise est moindre que $\frac{9,4}{2\times 4,12}(0,002)^2$, et à fortiori moindre que $\frac{10}{2\times 4}\cdot\frac{4}{10^6}$ ou $\frac{1}{2.10^5}$. Il y a en outre une seconde erreur provenant de ce que le dernier chiffre de la valeur $b_1=1,80195$ a été pris par excès. Cette erreur

est moindre que $\frac{1}{10^5}$, et de même sens que la précédente. Ainsi 1,80195 est approché par excès à $\frac{3}{2}$ cent-millièmes près; en sorte que la racine tombe entre 1,80193 et 1,80195. Donc, en définitive, on pourra prendre 1,80194 pour valeur approchée de la racine à 1 cent-millième près, et il restera à substituer ce nombre dans le premier membre de l'équation proposée pour fixer le sens de l'approximation. Le résultat de la substitution est positif + 0,0000093; par conséquent, 1,80194 est approché par excès.

88. Dans l'application que nous venons de faire, le calcul *d'un seul* terme de correction,

$$-\frac{f(1,8)}{f'(1,8)},$$

a fourni immédiatement cinq décimales exactes.

D'ordinaire l'approximation n'est pas aussi rapide. L'erreur commise ayant pour limite supérieure $\frac{M}{2N}h^2$, on peut dire qu'elle est en général de l'ordre du carré de h. Si donc, au début de la méthode de Newton, le calcul ne fournit pas sur h d'autre donnée que celle-ci : $h < \frac{1}{10}$, l'erreur sera *communément* (*) de l'ordre des centièmes; et il serait alors illusoire de pousser le calcul du premier terme de correction $-\frac{f(a)}{f'(a)}$ au delà de la seconde décimale.

Pareillement, si dans le calcul du second terme de correction $-\frac{f(a_1)}{f'(a_1)}$, on ne sait rien sur h, si ce n'est qu'il est

(*) Nous disons *communément*, car il peut se présenter tel cas où le facteur $\frac{M}{2N}$ serait beaucoup au-dessous de 1, par exemple plus petit que $\frac{1}{10}$; alors la limite $\frac{M}{2N}h^2$ tomberait au-dessous de $\frac{1}{100}$.

plus petit que $\frac{1}{100}$, il est clair qu'on ne prendra pas plus de quatre décimales au quotient, et ainsi de suite en doublant.

Au contraire, dans le calcul ci-dessus, le quotient $-\frac{f(1,8)}{f'(1,8)}$ nous ayant fourni une valeur de *h par excès et plus petite que* $\frac{2}{1000}$, il était logique de pousser immédiatement ce quotient jusqu'à la cinquième décimale.

89. Voici un autre exemple :

On propose de résoudre l'équation

$$x^3 + x - 1000 = 0.$$

Le coefficient de x étant positif et celui de x^2 étant nul, on sait que cette équation n'a qu'une seule racine réelle. L'hypothèse $x = 10$ réduit le premier membre à $+10$, et $x = 9$ donne -262. Ainsi la racine est comprise entre 9 et 10, mais beaucoup plus voisine de 10 que de 9. Posons

$$f(x) = x^3 + x - 1000, \quad \text{d'où} \quad f'(x) = 3x^2 + 1, \quad f''(x) = 6x.$$

Soit $b = 10$; $f(b)$ et $f''(b)$ étant de même signe, le terme de correction

$$-\frac{f(b)}{f'(b)}$$

fournira une approximation de même sens que b, ou par excès. On trouve.

$$-\frac{f(b)}{f'(b)} = -\frac{10}{301} = -0,033\ldots.$$

A quelle décimale convient-il de s'arrêter? Si l'on évalue, une fois pour toutes, la fraction $\frac{M}{2N}$ qui est très-simple, on

trouve

$$\frac{M}{2N} = \frac{f''(10)}{2f'(9)} = \frac{60}{3.244} = \frac{15}{122} < 0,2.$$

D'ailleurs, bien que la valeur calculée de $h = 0,033\ldots$ soit par défaut, il est bien évident que la partie négligée ne saurait influer sur les dixièmes de cette valeur, en sorte que l'on a $h < 0,1$; donc

$$\frac{M}{2N} h^2 < 0,002.$$

On est ainsi conduit à prendre trois décimales au quotient $\frac{f(b)}{f'(b)}$, soit 0,033; et la valeur approchée qui en résulte pour la racine,

$$b_1 = 10 - 0,033 = 9,967,$$

n'est fautive par excès que de *deux* unités du dernier ordre, *au plus*.

Si l'on veut s'arrêter là dans l'approximation, il reste à vérifier le dernier chiffre décimal. A cet effet on n'a qu'à substituer dans l'équation 9,966 à la place de x. Il vient

$$f(9,966) = -0,199\ldots,$$

résultat dont le signe nous apprend que b_1 est approchée par excès à moins de $\frac{1}{1000}$. (Comme le *signe* du résultat importe seul, ce calcul peut s'abréger à l'aide des Tables de Callet, bien qu'elles ne donnent que les trois premières décimales du cube de 9,966).

Mais, si l'on doit poursuivre l'approximation, cette vérification n'est pas nécessaire, et l'on peut immédiatement procéder au calcul du second terme de correction,

$$-\frac{f(b_1)}{f'(b_1)} = -\frac{0,099634063}{299,023267}.$$

On voit que les trois premiers chiffres du quotient seront

des zéros; et vu la petitesse de $\frac{M}{2N}$, on est conduit à pousser cette fois la division jusqu'à la *septième* décimale. Il y aura donc *quatre* chiffres significatifs à chercher. Il suffit pour cela (29) de conserver six chiffres au dividende et au diviseur (à partir du premier significatif). On aura soin de forcer l'unité du dernier chiffre conservé au diviseur, afin que le quotient ne cesse pas d'être par défaut. Ainsi l'on divisera simplement 996340 par 299024 par la règle de la division abrégée. Il vient

$$-\frac{f(b_1)}{f'(b_1)} = -0{,}0003331 \text{ (par défaut)},$$

d'où

$$b_2 = 9{,}967 - 0{,}0003331 = 9{,}9666669.$$

Cette valeur est toujours par excès, mais le dernier chiffre 9 n'est pas sûr; il peut être fautif d'une ou deux unités au plus. Pour le vérifier, on substitue dans l'équation 9,9666668 à la place de x. Le résultat est positif, mais si peu supérieur à zéro, qu'on aperçoit clairement, sans recommencer le calcul du cube, que si l'on ôtait encore une unité du dernier ordre, c'est-à-dire si l'on substituait 9,9666667 on aurait un résultat négatif.

En définitive,

$$9{,}9666668$$

est une valeur approchée par excès de la racine à moins de $\frac{1}{10^7}$.

En employant la méthode de Newton, avec les précautions que nous venons d'exposer, on voit que l'on marche sûrement vers la racine cherchée par une série de termes calculés chacun avec le nombre de décimales qui lui est propre, et qui fournissent une approximation croissante et toujours de même sens. Une *seule* vérification finale est

nécessaire pour reconnaître si le dernier chiffre décimal est exact, ou s'il doit être haussé ou baissé de quelques unités.

90. *Exercices.* — I. L'équation

$$x^3 - 3x + 1 = 0$$

a une racine comprise entre 1,5 et 1,6. En partant de 1,6, la méthode de Newton fournira une suite d'approximations par excès : on trouvera

$$x = 1,53208\ldots$$

II. L'équation

$$5x^3 - 27x^2 - 180x - 324 = 0$$

n'a qu'une seule racine réelle et positive comprise entre 9 et 10. En substituant des nombres équidistants de $\frac{1}{10}$, on trouve que 9,7 et 9,8 donnent des résultats de signes contraires. D'ailleurs le premier membre de l'équation et sa dérivée seconde ont le même signe + quand on y fait $x = 9,8$ On partira donc de ce nombre pour appliquer la méthode. Le calcul du premier terme de correction fournira trois décimales, la première étant zéro; et le calcul du second terme en donnera sept. La dernière division s'abrégera par l'emploi des logarithmes. On trouvera

$$x = 9,7657913 \quad \text{(par excès).}$$

91. Toutes les fois que l'équation $f(x) = 0$ est *algébrique*, et aussi *dans certains cas d'équations transcendantes*, on peut s'affranchir de toute hypothèse sur le signe de la dérivée seconde $f''(x)$.

Soient, en effet, $P(x)$ le groupe des termes positifs de la première dérivée $f'(x)$, et $-N(x)$ le groupe des termes négatifs de cette même dérivée, en sorte que

$$f'(x) = P(x) - N(x).$$

Supposons, comme au n° 78, $f(a) < 0$, et par suite

$f'(a+\theta h)>0$. Comme chacun des polynômes algébriques $P(x)$, $N(x)$ va croissant avec x, si l'on fait $x=b$ dans le premier et $x=a$ dans le second, il viendra

$$f'(a+\theta h)<P(b)-N(a).$$

L'équation (9) donnera ensuite

$$h>\frac{-f(a)}{P(b)-N(a)}.$$

Cette limite de h conviendra évidemment encore à une équation transcendante, pourvu que $P(x)$ et $N(x)$ soient des fonctions positives croissantes.

Soit donc

$$a_1=a-\frac{f(a)}{P(b)-N(a)};$$

a_1 sera une seconde valeur plus approchée que a de la racine, et dans le même sens, c'est-à-dire par défaut.

Ceci suppose $f(a)<0$: si l'on avait $f(a)>0$, et, par conséquent, $f'(a+\theta h)<0$, il faudrait d'abord renverser les signes de $f'(x)$ et écrire

$$-f'(x)=N(x)-P(x),$$

d'où l'on tirerait

$$-f'(a+\theta h)<N(b)-P(a);$$

donc, en transportant dans l'équation (9) le signe — au dénominateur, et remplaçant la quantité positive $-f'(a+\theta h)$ par $N(b)-P(a)$, il viendrait

$$h>\frac{f(a)}{N(b)-P(a)} \quad \text{ou} \quad h>-\frac{f(a)}{P(a)-N(b)}.$$

Enfin

$$a-\frac{f(a)}{P(a)-N(b)}$$

remplacerait l'expression donnée plus haut pour a_1. On voit qu'elles ne diffèrent l'une de l'autre que par un échange entre les nombres a et b dans le dénominateur.

Continuons à raisonner dans la première hypothèse où l'on a $f(a) < 0$. Si nous partons maintenant de la valeur par excès b, et que nous désignions par k ce qu'il faut en retrancher pour avoir la racine exacte, l'équation (9) dans laquelle nous remplacerons h par $-k$ et a par b donnera

$$k = \frac{f(b)}{f'(b - \theta k)}.$$

Comme on a $f(b) > 0$ et $k > 0$, on a aussi $f'(b - \theta k) > 0$; et d'ailleurs

$$f'(b - \theta k) < \mathrm{P}(b) - \mathrm{N}(a),$$

d'où

$$k > \frac{f(b)}{\mathrm{P}(b) - \mathrm{N}(a)}.$$

Soit donc

$$b_1 = b - \frac{f(b)}{\mathrm{P}(b) - \mathrm{N}(a)};$$

b_1 sera une nouvelle valeur plus approchée que b de la racine, et dans le même sens, c'est-à-dire par excès.

Comme on calcule d'ordinaire a_1 et b_1 en décimales, on s'arrêtera pour a_1 à un nombre par défaut, et pour b_1 à un nombre par excès.

L'erreur commise en adoptant a_1 ou b_1 pour valeur de la racine, sera moindre que la différence $b_1 - a_1$. Les décimales communes à ces deux nombres appartiendront à l'expression exacte de la racine.

Pour continuer l'approximation, on répétera sur a_1 et b_1 les mêmes calculs que sur a et b, et l'on parviendra ainsi à deux nouvelles limites a_2 et b_2 plus rapprochées de la racine, et ainsi de suite.

Ces expressions de a_1, b_1 sont dues à M. Cauchy. Elles conviennent à tous les cas que peut offrir la résolution d'une équation *algébrique ;* cependant elles ne fournissent pas toujours une approximation aussi rapide que la méthode de Newton, rectifiée comme on l'a vu aux n°s 78 et 79.

On peut s'en convaincre en les appliquant à l'équation numérique déjà traitée (80).

92. Nous allons maintenant appliquer la méthode de Newton et celle des parties proportionnelles à la résolution de quelques équations transcendantes.

Proposons-nous d'abord de *calculer la valeur approchée du plus petit arc de cercle qui a la propriété d'être égal à sa tangente trigonométrique.* (On exclut l'arc o degré.)

Ce calcul se présente dans le problème dont voici l'énoncé : *Parmi tous les arcs de cercle d'une même longueur donnée et de rayons variables, déterminer celui pour lequel l'aire du segment correspondant est maxima* (*).

En désignant par l la longueur de l'arc ACB (*fig.* 13) et par r

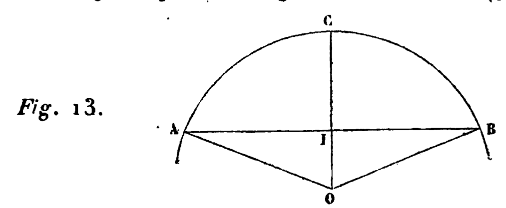


Fig. 13.

le rayon OA, on a

$$\text{segment ACBI} = \frac{1}{2}\left[lr - r^2 \sin \frac{l}{r}\right].$$

Si l'on égale à zéro la dérivée de cette fonction de r, on trouve une équation qui se partage en deux, savoir :

$$\cos \frac{l}{2r} = 0 \quad \text{et} \quad \text{tang}\, \frac{l}{2r} = \frac{l}{2r}.$$

La première donne

$$r = \frac{l}{(2n+1)\pi},$$

d'où résultent une infinité de segments *maxima* qu'on obtient en faisant croître l'entier n de o à ∞. Parmi eux le *maximum maxi-*

(*) Voir mon *Cours complémentaire d'Analyse*, page 132.

morum répond à $n = 0$, d'où $r = \frac{l}{\pi}$: ce segment est évidemment un demi-cercle.

La seconde équation devient, en posant $\frac{l}{2r} = x$,

$$(13) \qquad \operatorname{tang} x - x = 0.$$

Elle admet une infinité de racines réelles, égales et de signes contraires deux à deux, que l'on peut construire par l'intersection de la droite $y = x$ avec la courbe à branches infinies $y = \operatorname{tang} x$. Si l'on désigne par α, β, γ, etc., la suite des racines positives, rangées par ordre de grandeur (excepté la racine zéro), la construction graphique permet de les séparer aisément ; on a

$$\pi < \alpha < \frac{3\pi}{2}, \quad 2\pi < \beta < \frac{5\pi}{2}, \quad 3\pi < \gamma < \frac{7\pi}{2} \cdots$$

A ces racines répondent des segments *minima*. Ces segments s'intercalent entre les maxima obtenus plus haut. Car les valeurs de $\frac{l}{2r}$ qui répondent aux maxima sont $\frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}$, etc. Les maxima et minima d'une fonction doivent en effet se succéder alternativement.

93. Laissant de côté toute discussion de maximum et de minimum, nous nous proposons de calculer approximativement la racine α de l'équation (13).

Deux modes de calcul se présentent : l'un consiste à remplacer l'équation proposée par l'*équation logarithmique*,

$$(14) \qquad \log \operatorname{tang} x - \log x = 0;$$

puis à employer la *méthode des parties proportionnelles*, après avoir resserré convenablement les limites entre lesquelles la racine est comprise. Dans l'autre mode, on laisse l'équation (13) telle qu'elle est, afin de ne pas compliquer

les calculs à effectuer sur la dérivée, et l'on applique la *Méthode de Newton.*

Nous suivrons d'abord le premier mode.

Prenons donc l'équation (14). La première chose à faire, est de rapprocher les limites entre lesquelles on sait qu'est comprise la racine : ces limites sont les arcs de 180° et 270°. Pour cela, nous substituerons d'abord à la place de x une suite de moyennes arithmétiques.

Posons

$$f(x) = \log \operatorname{tang} x - \log x :$$

1°. $x = \text{arc}\ (180^\circ + 45^\circ)$ ou 225° donne

$$f(225^\circ) = \log \operatorname{tang} 45^\circ - \log(\text{arc}\, 225^\circ) = 0 - \log(\text{arc}\, 225^\circ),$$

résultat évidemment négatif. L'arc cherché est donc compris entre 225° et 270°.

2°. Faisons $x = \text{arc}\ (180^\circ + 67^\circ\, 30')$, ou simplement, en négligeant les minutes, $x = \text{arc}\ (180^\circ + 67^\circ)$ ou 247°.

Ici, et dans les essais qui vont suivre, nous aurons besoin des valeurs numériques des arcs rapportées au rayon pris pour unité. Ces valeurs se trouvent toutes calculées dans une petite table qui occupe trois pages des Tables de Callet, immédiatement *avant* les logarithmes-sinus relatifs à la division centésimale du cercle. Cette petite table, intitulée : *Rapports des longueurs des degrés au rayon*, donne non-seulement les valeurs des degrés depuis 1° jusqu'à 100°, mais les minutes de 1′ à 60′, et les secondes de 1″ à 100″. — D'après cela, pour avoir la longueur de l'arc 247°, on n'aura que cette simple addition à faire :

$$\begin{array}{l} \text{arc } 180^\circ = 3,1415926 \\ \text{arc } \;\;67^\circ = 1,1693706 \text{ (donné par la Table)} \\ \hline \text{arc } 247^\circ = 4,3109632 \end{array}$$

Il reste à prendre le logarithme de cet arc dans la Table des nombres, et l'on a

$$\log(\text{arc } 247^\circ) = 0,6345743.$$

C'est ainsi qu'on opérera dans tous les essais suivants. Remarquons encore que pour avoir les logarithmes des arcs avec sept décimales exactes, il convient (60) de prendre leurs rapports au rayon avec *huit* chiffres. C'est ce qu'on a fait ci-dessus.

Cela posé, on aura, pour $x = 247^\circ$,

$$\begin{aligned} f(247^\circ) &= \log\tan 67^\circ - \log(\text{arc } 247^\circ) \\ &= 0,3721481 = 0,6345743, \end{aligned}$$

résultat négatif. La racine est donc comprise entre 247° et 270°.

3°. $x = 180^\circ + 78^\circ$ ou 258° donne

$$\begin{aligned} f(258^\circ) &= \log\tan 78^\circ - \log(\text{arc } 258^\circ) \\ &= 0,6725255 - 0,6534971 = +0,0190284, \end{aligned}$$

résultat positif. Arc 258° est donc une limite supérieure de la racine.

Nous avons déjà pour limite inférieure 247°; mais, vu les résultats qu'ont donnés les substitutions précédentes et la rapidité avec laquelle croît $\tan x$, il y a lieu de penser que la racine est beaucoup plus voisine de 258° que de 247°. C'est pourquoi, abandonnant les moyennes arithmétiques, essayons 257°. Il vient

$$\begin{aligned} f(257^\circ) &= \log\tan 77^\circ - \log(\text{arc } 257^\circ) \\ &= 0,6366359 - 0,6518105 = -0,0151746. \end{aligned}$$

Il y a changement de signe. La racine est donc comprise entre 257° et 258° : elle est connue à 1° près.

Posons

$$a = 257°, \quad b = 258° \ (*).$$

Pour savoir de laquelle de ces deux valeurs il convient de partir en employant la méthode des parties proportionnelles, examinons les signes des dérivées première et seconde de la fonction $f(x)$. Il vient

$$f'(x) = M\left(\frac{2}{\sin 2x} - \frac{1}{x}\right),$$

$$f''(x) = M\left(\frac{-4\cos 2x}{\sin^2 2x} + \frac{1}{x^2}\right);$$

M désigne le module $\frac{1}{l\,10}$. Comme x est compris entre π et $\frac{3\pi}{2}$, $2x$ est compris entre 2π et 3π, et, par conséquent, $\sin 2x$ est positif; d'ailleurs on a $2x > \sin 2x$. Donc aussi

$$\frac{2}{\sin 2x} > \frac{1}{x};$$

$f'(x)$ reste donc positive pour toutes les valeurs de x comprises entre a et b.

Quant à $f''(x)$, il est évident que cette dérivée reste aussi positive, puisque $\cos 2x$ est négatif.

Enfin pour la valeur $x = a$, $f(x)$ et $f''(x)$ sont de signes contraires.

(*) Si l'on faisait une *troisième substitution*, celle de 259°, on aurait les *deux différences premières*

$$\Delta_1 = f(258°) - f(257°), \quad \Delta_2 = f(259°) - f(258°),$$

correspondantes à des arcs équidistants de 1°; d'où l'on pourrait conclure *la différence seconde*

$$\Delta_2 - \Delta_1,$$

et l'on reconnaîtrait que cette dernière est une très-petite fraction vis-à-vis des différences premières, et, par conséquent, négligeable dans le premier degré d'approximation. Ceci légitimerait au besoin, à priori, la méthode des parties proportionnelles qui revient, comme on sait, à considérer les différences secondes de la fonction comme nulles.

Donc (84), en partant de $a = 257^\circ$, on sera sûr que le terme de correction

$$-\frac{(b-a)f(a)}{f(b)-f(a)}$$

fournira une approximation croissante et de même sens que a.

Voici le tableau des calculs.

Première approximation. — On a

$$b - a = 1^\circ = 60',$$
$$f(a) = -0{,}0151746,$$
$$f(b) = +0{,}0190284,$$

d'où

$$-\frac{(b-a)f(a)}{f(b)-f(a)} = 60' \cdot \frac{151746}{342030} = \frac{910476'}{34203}.$$

Par suite de la multiplication par 6, le dernier chiffre du dividende peut être fautif de plusieurs unités; mais cela est indifférent, attendu que, dans cette première approximation, nous ne conserverons que le nombre entier de minutes de l'arc. Nous n'avons donc qu'à évaluer ce quotient à un dixième près; et, pour cela, il suffit (29) de conserver quatre chiffres tant au dividende qu'au diviseur, en ayant soin de forcer l'unité sur ce dernier, comme l'indique le tableau ci-dessous :

9104	3421
22620	26',6.
2094	

Ajoutant 26' à 257°, on aurait une approximation par défaut de la racine; mais, eu égard à la fraction de minute qui surpasse $\frac{1}{2}$, il vaut mieux adopter 257° 27', valeur qui sera *probablement* encore approchée par défaut. On s'en assure en la substituant dans l'équation : cette substitu-

tion servira en tout cas pour la suite de l'approximation.

$$\begin{aligned} f(257^\circ 27') &= \log \operatorname{tang}(77^\circ 27' - \log(\operatorname{arc} 257^\circ 27') \\ &= 0,6524546 - 0,6525703 \\ &= -0,0001157. \end{aligned}$$

Le signe — nous apprend que $257^\circ 27'$ est en effet approché par défaut; et la petitesse du résultat nous conduit en même temps à essayer $257^\circ 28'$, afin d'abaisser, s'il se peut, la limite supérieure de la racine.

Il vient

$$\begin{aligned} f(257^\circ 28') &= \log \operatorname{tang}(77^\circ 28') - \log(\operatorname{arc} 257^\circ 28') \\ &= 0,6530506 - 0,6525984 \\ &= +0,0004522, \end{aligned}$$

résultat positif. Ainsi $257^\circ 27'$ est approché à $1'$ près et par défaut.

Deuxième approximation. — On posera maintenant

$$a_1 = 257^\circ 27', \quad b_1 = 257^\circ 28';$$

d'où

$$\begin{aligned} b_1 - a_1 &= 1' = 60'', \\ f(a_1) &= -0,0001157, \\ f(b_1) &= +0,0004522, \end{aligned}$$

$$-\frac{(b_1 - a_1) f(a_1)}{f(b_1) - f(a_1)} = \frac{60''.1157}{5679} = \frac{69420''}{5679} = 12'',2\ldots.$$

Ajoutant ce nombre à $257^\circ 27'$, il vient pour seconde approximation $257^\circ 27' 12'',2$.

On s'assure par substitution que $257^\circ 27' 12'',3$ donne un résultat positif; d'où l'on conclut que

$$a_2 = 257^\circ 27' 12'',2$$

est une valeur approchée de la racine, par défaut et à moins de $\frac{1}{10}$ de seconde.

94. Nous allons maintenant indiquer rapidement comment on conduirait le calcul en suivant le second mode, c'est-à-dire la *méthode de Newton*.

On laissera, comme nous l'avons déjà dit, l'équation sous la forme

$$(13) \qquad \tan x - x = 0.$$

Puis, par les mêmes essais que précédemment, on arrivera à renfermer la racine entre les limites 257° et 258°. Seulement, au lieu de chercher le logarithme de l'arc x pour le retrancher de celui de tang x, on aura à remonter du logarithme de tang x à tang x, pour en retrancher le nombre x que donne immédiatement la petite table.

Soit $f(x) = \text{tang}\, x - x$.

Pour $x = 257°$, on aura

$$f(257°) = \text{tang}\, 77° - \text{arc}\, 257°$$
$$= 4,331475 - 4,485496 = -0,154021.$$

Maintenant, pour connaître le sens de l'approximation que fournira la méthode, appliquée à la valeur $a = 257°$, on examinera les signes des dérivées

$$f'(x) = \frac{1}{\cos^2 x} - 1 = \text{tang}^2 x,$$

et

$$f''(x) = \frac{2\, \text{tang}\, x}{\cos^2 x},$$

qui sont plus simples que dans le premier mode. Ces dérivées restent évidemment positives entre 257° et 258°; et de plus, pour $x = 257°$, $f(x)$ et $f''(x)$ sont de signes contraires. Donc (83), en partant de la valeur $a = 257°$, l'approximation fournie par le terme de correction

$$-\frac{f(a)}{f'(a)}$$

changera de sens, et fournira une valeur en excès.

Suivons néanmoins cette marche. Afin que les résultats du calcul soient immédiatement exprimés en secondes, nous multiplierons $\frac{f(a)}{f'(a)}$ par $\frac{1}{\sin 1''}$ (*) et il viendra

$$h < \frac{-f(a)}{f'(a)} \cdot \frac{1}{\sin 1''};$$

h désigne actuellement le nombre de secondes qu'il faut ajouter à a pour avoir la racine cherchée. Le calcul se fera par logarithmes.

Première approximation. — On a

$$\log \frac{1}{\sin 1''} = 5,3144251,$$

$$\log[-f(a)] = \bar{1},1875800 \text{ (par excès)},$$

$$\log f'(a) = 2 \log \operatorname{tang} 77^\circ = 1,2732718.$$

Ajoutant les deux premiers logarithmes et retranchant le troisième, il vient

$$\log h < 3,2287333, \quad \text{d'où} \quad h < 1694'' = 28'\,14''.$$

Ce nombre ajouté à 257° donnerait une première approximation par excès, $257^\circ\,28'\,14''$.

Mais, comme on ne peut pas espérer que les secondes soient déjà exactes et que leur nombre est petit, on les négligera et on adoptera la valeur

$$b_1 = 257^\circ\,28',$$

laquelle sera *probablement* encore approchée par excès.

(*) En général, soit l la longueur d'un arc rapportée au rayon, h le nombre de secondes de cet arc, on a

$$l = h.\text{arc}\,1'' \quad \text{ou sensiblement} \quad h \sin 1'';$$

d'où

$$h = l\,\frac{1}{\sin 1''}.$$

On sait d'ailleurs que l'erreur commise en prenant $\sin 1''$ pour $\text{arc}\,1''$ est moindre que $\frac{1}{4.10^{16}}$.

On s'en assurera par substitution. Il vient en effet

$$f(b_1) = +0,004681.$$

Deuxième approximation. — Le terme de correction

$$\frac{f(b_1)}{f'(b_1)} \cdot \frac{1}{\sin 1''},$$

retranché de b_1, fournira (78) une approximation de même sens que b_1, c'est-à-dire par excès.

$$\log \frac{1}{\sin 1''} = 5,3144251,$$

$$\log f(b_1) = 3,6703386 \text{ (par défaut)},$$

$$\log f'(b_1) = 1,3061012.$$

Si l'on désigne par k le nombre de secondes qu'il faut retrancher de b_1 pour avoir la racine, on aura

$$\log k > 1,6786625, \quad k > 47'',7.$$

D'où l'on conclut la seconde valeur approchée par excès

$$b_2 = 257^\circ\, 27'\, 12'',3.$$

Cette valeur n'est pas fautive de $\frac{1}{10}$ de seconde. En effet, le premier mode de calcul avait donné pour valeur approchée *par défaut*, $a_2 = 257^\circ\, 27'\, 12'',2$.

Remarque. — Dans la question que nous venons de résoudre, on voit que les deux modes de calcul ont à peu près également bien réussi. Mais on comprend que la méthode de Newton, qui exige le retour aux nombres, n'aurait pas le même succès dans un autre cas où la ligne trigonométrique serait combinée, non plus avec x (première puissance), mais avec $\frac{1}{x}$, $\frac{1}{x^2}$ par exemple. L'emploi de l'é-

quation logarithmique et des parties proportionnelles serait alors plus simple.

95. Comme exercice, on peut se proposer de calculer directement, par l'équation (12) du n° 86, l'erreur que comporte la valeur approchée b_2. Pour que cette erreur soit immédiatement traduite en fraction de seconde, nous remplacerons dans l'équation (12) h par $-k \sin 1''$, et d'ailleurs a par b_1. Puis, résolvant par rapport à k, il viendra

$$k = \frac{f(b_1)}{f'(b_1)} \cdot \frac{1}{\sin 1''} + \frac{k^2}{1.2} \sin 1'' \cdot \frac{f''(b_1 - \theta k \sin 1'')}{f'(b_1)}.$$

Ainsi, en désignant l'erreur cherchée par ε, on a

$$\varepsilon = \frac{k^2}{1.2} \sin 1'' \cdot \frac{f''(b_1 - \theta k \sin 1'')}{f'(b_1)};$$

et, comme $f''(x)$ est une fonction croissante,

$$\varepsilon < \frac{k^2}{1.2} \sin 1'' \cdot \frac{f''(b_1)}{f'(b_1)} \quad \text{ou} \quad \frac{2k^2 \sin 1''}{\sin 2b_1}.$$

Ici l'on ne peut pas mettre pour k la valeur calculée $47'',7$, puisqu'elle est par défaut; mais bien une valeur en excès, par exemple $1'$ ou $60''$. D'ailleurs $b_1 = 257^\circ 28'$; il vient donc

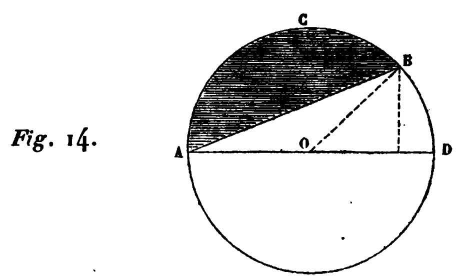
$$\varepsilon < \frac{2(60)^2 \sin 1''}{\sin(25^\circ 4')} = \frac{7200 \sin 1''}{\sin(25^\circ 4')};$$

$$\log \varepsilon < \bar{2},9158771.$$

La caractéristique étant -2, on voit, sans remonter au nombre, que ε sera moindre que $\frac{1}{10}$ de seconde, ce qui s'accorde bien avec les vérifications précédentes.

96. Problème. — *Par un point donné* A *sur la circonférence d'un cercle, on propose de tirer une corde* AB *qui retranche un segment* ACB *équivalent au quart de l'aire du cercle* (*).

Fig. 14.



Prenons pour unité le rayon AO (*fig.* 14); pour inconnue l'arc ACB que nous désignerons par s. On a

$$\text{segment ACB} = \frac{1}{2}s - \frac{1}{2}\sin s;$$

et, par conséquent, l'équation du problème est

$$s - \sin s = \frac{\pi}{2};$$

ou bien, en posant $s - \frac{\pi}{2} = x$,

(15) $$x - \cos x = 0.$$

Ainsi la question revient à *trouver un arc égal à son cosinus.*

La fonction $x - \cos x$ est croissante : or, quand x croît de 0 à $\frac{\pi}{2}$, cette fonction passe de la valeur -1 à $+\frac{\pi}{2}$. Donc l'équation (15) a une racine réelle et une seule, comprise entre 0 et $\frac{\pi}{2}$. Il faut rapprocher ces limites. Nous rempla-

(*) Euler, *Introduction à l'Analyse infinitésimale*, t. II, p. 316.

cerons l'équation (15) par l'équation logarithmique

$$(16)\qquad \log x - \log\cos x = 0,$$

et nous poserons

$$f(x) = \log x - \log\cos x.$$

1°. Pour $x = 45°$ ou $\frac{\pi}{4}$, il vient

$$f(45°) = \bar{1},895\ldots, \quad -\bar{1}.849\ldots, \quad > 0;$$

le signe de cette différence nous apprend que x est inférieur à 45°.

2°. $x = 40°$ donne

$$f(40°) = \bar{1},8439374 - \bar{1},8842540 = -0,0403166.$$

La racine est donc comprise entre 40° et 45°.

Comme la fonction $\log x - \log\cos x$ varie lentement, ces limites sont assez rapprochées pour que l'on puisse leur appliquer la méthode des parties proportionnelles. C'est ainsi que procède Euler. Il y a cependant ici une remarque à faire.

Si l'on prend les dérivées de $f(x)$, on trouve

$$f'(x) = M\left(\frac{1}{x} + \operatorname{tang} x\right),$$

$$f''(x) = M\left(\frac{1}{\cos^2 x} - \frac{1}{x^2}\right), \quad \left(M = \frac{1}{l\,10}\right).$$

$f'(x)$ reste évidemment positive entre 40° et 45°; mais $f''(x)$ change de signe en passant par zéro, précisément pour la valeur cherchée de x. Ainsi la courbe définie par l'équa-

tion

$$y = \log x - \log \cos x$$

présente une inflexion au point R, où elle coupe l'axe des x, comme le montre la *fig.* 15. De concave qu'elle était de m

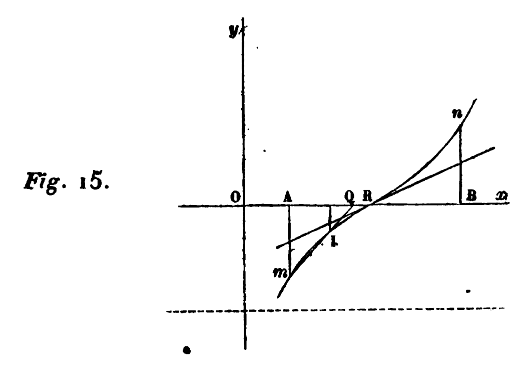


Fig. 15.

jusqu'à R, vers une parallèle inférieure à l'axe des x, elle devient convexe de R jusqu'à n. A cause de cette particularité, si l'on applique la méthode des parties proportionnelles en partant des valeurs 40° et 45°, ce qui revient géométriquement à remplacer la courbe par la sécante mn, le sens de l'approximation restera incertain; car on ne prévoit pas si cette transversale ira couper l'axe des x en deçà ou au delà de R.

Ce sera donc ici le cas de modifier le procédé, comme on l'a dit au n° 81. En partant de deux valeurs de x *par défaut*, telles que celles qui répondent aux deux points m et I, on sera sûr que le terme de correction AQ donnera une approximation par défaut.

A cet effet, essayons 41°. Il vient

$$\begin{aligned} f(41^\circ) &= \bar{1},8546613 - \bar{1},8777799 \\ &= -0,0231186; \end{aligned}$$

ainsi l'arc 41° est encore inférieur à la racine, et peut con-

séquemment être associé à l'arc 40°. Nous ferons donc

$$a = 40^\circ, \quad b = 41^\circ \ (*);$$

d'où

$$b - a = 1^\circ, \quad f(a) = -0{,}0403166, \quad f(b) = -0{,}0231186,$$

$$-\frac{(b-a)f(a)}{f(b)-f(a)} = \frac{403166^\circ}{171980} = 2^\circ\,20' \quad \text{(par défaut)}.$$

En s'arrêtant aux minutes dans ce premier calcul, on a $42^\circ\,20'$ pour valeur approchée par défaut de la racine.

Pour continuer l'approximation, on substituera $42^\circ\,20'$ dans l'équation (16)

$$f(42^\circ\,20') = \bar{1}{,}8685599 - \bar{1}{,}8687851$$
$$= -0{,}0002252.$$

Puis, en posant

$$a_1 = 41^\circ, \quad b_1 = 42^\circ\,20',$$

on aura

$$b_1 - a_1 = 80', \quad f(a_1) = -0{,}0231186,$$
$$f(b_1) = -0{,}000222;$$

$$-\frac{(b_1 - a_1)f(a_1)}{f(b_1) - f(a_1)} = \frac{80'.231186}{228934} = 80',7869 = 1^\circ\,20'\,47'',2.$$

Ainsi on a, pour seconde approximation,

$$a_2 = 42^\circ\,20'\,47'',2.$$

On s'assure que cette valeur est approchée par défaut à moins de $\frac{1}{10}$ de seconde.

Telle est valeur de l'arc égal à son cosinus. Maintenant,

(*) Si l'on fait encore la substitution de 42° dans l'équation, on pourra s'assurer, en formant les *deux différences premières*

$$\Delta_1 = f(41^\circ) - f(40^\circ), \quad \Delta_2 = f(42^\circ) - f(41^\circ),$$

puis la *différence seconde* $\Delta_2 - \Delta_1$, que cette dernière est une très-petite fraction comparée aux différences premières. Elle est moindre que 0,0003.

pour revenir au problème proposé, on a $s = \frac{\pi}{2} + x$, d'où, en degrés,

$$s = 132^\circ\, 20'\, 47'',2.$$

La corde qui sous-tendra ce dernier arc retranchera du cercle un segment équivalent au quart de sa surface.

Euler parvient au résultat que nous venons d'obtenir à l'aide de *trois* proportions successives, en partant : 1° des limites 40° et 45°, ce qui lui donne les 2° de l'arc cherché ; 2° des limites 42° et 43°, ce qui donne les 20′ ; 3° des limites 42° 20′ et 42° 21′, ce qui fournit les 47″,2. Mais le sens de l'approximation reste incertain jusqu'après vérification, et puis il y a une proportion de plus à calculer que dans notre procédé.

Si l'équation du problème avait été conservée sous la forme

$$(15) \qquad x - \cos x = 0,$$

on aurait pu lui appliquer aussi avec succès la méthode de Newton. La particularité relative au point d'inflexion, que nous avons signalée plus haut, ne se serait plus présentée.

97. Problème. — *Déterminer sur une circonférence de cercle un arc* AmB, *dont la longueur soit égale à la somme faite du sinus* BP *et de la distance* AP (*fig.* 16).

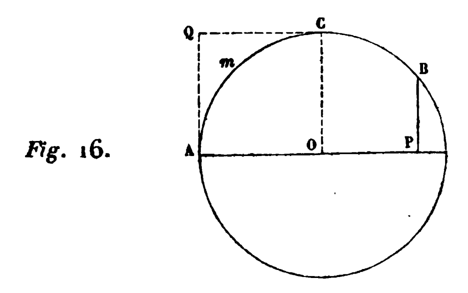


Fig. 16.

En désignant l'arc AmB par s et continuant à prendre le

rayon pour unité, on a l'équation

$$s = \sin s + 1 - \cos s.$$

Pour l'approprier au calcul logarithmique, on la transforme ainsi :

$$s = 2 \sin \frac{1}{2} s \left(\cos \frac{1}{2} s + \sin \frac{1}{2} s \right)$$

$$= 2 \sqrt{2} \sin \frac{1}{2} s \cos \left(\frac{1}{2} s - 45^\circ \right),$$

et, si l'on pose $\frac{s}{2} = x$, il vient

$$x - \sqrt{2} \sin x \cos(x - 45^\circ) = 0;$$

ou mieux, pour le calcul,

$$(17) \quad \log x - \left[\frac{\log 2}{2} + \log \sin x + \log \cos(x - 45^\circ) \right] = 0.$$

Comme l'arc de 90° ou AmC est évidemment plus petit que la somme des lignes enveloppantes $AQ + QC$ ou $AO + OC$, on en conclut que s doit surpasser 90°, et par suite x est plus grand que 45° ; x doit même être assez éloigné de cette limite.

L'hypothèse $x = 70^\circ$, faite dans l'équation (17), donne un résultat très-approchant de zéro et positif, et $x = 69^\circ$ donne un résultat de signe contraire. Ainsi la racine est comprise entre 69° et 70°.

La méthode des parties proportionnelles fournira, par deux proportions successives,

$$x = 69^\circ\, 5'\, 56'',5,$$

d'où

$$s \text{ ou l'arc } AmB = 138^\circ\, 11'\, 53''.$$

98. Problème. — *Déterminer, sur une circonférence de cercle, un arc* AmB *dont la longueur soit égale à la somme des tangentes* AT, BT, *menées par ses extrémités et terminées à leur point de concours* (*fig.* 17).

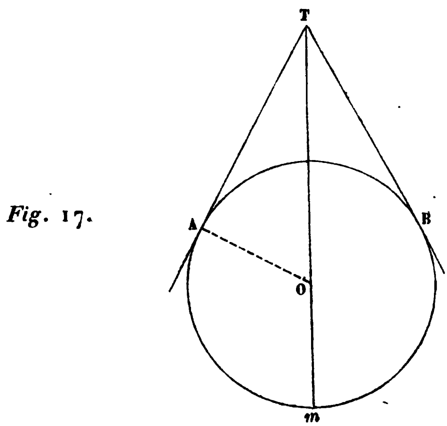


Fig. 17.

En posant arc $AmB = 2x$ et $OA = 1$, on a l'équation

$$x - \text{tang}(\pi - x) = 0.$$

Il est visible que x est compris entre 90° et 120°, et plus voisin de ce dernier nombre. En essayant $x = 115°$ dans le premier membre de l'équation, on obtient un résultat négatif. Il en est de même pour $x = 116°$, tandis que la substitution de 117° donne un résultat positif. La méthode de Newton appliquée à cette équation, ou celle des parties proportionnelles appliquée à l'équation logarithmique

$$\log x - \log \text{tang}(\pi - x) = 0,$$

donnera la racine

$$x = 116° \, 14' \, 21'',3 \text{ (par défaut).}$$

Par conséquent, l'arc cherché est

$$\text{arc } AmB = 232° \, 28' \, 42'' \text{ (à 1'' près).}$$

99. Problème. — *Déterminer sur une circonférence de cercle, un arc* AmB *tel, que l'aire du segment* AmB *soit égale à celle du triangle* AOB (*) (*fig.* 18).

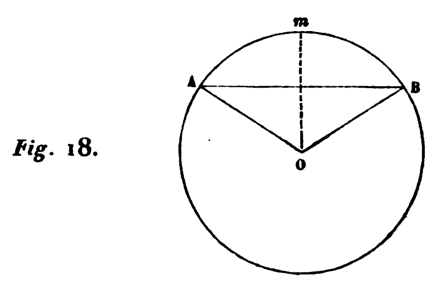


Fig. 18.

En posant arc $AmB = x$, $OA = 1$, on a l'équation

(18) $$x = 2 \sin x.$$

Le rapport $\frac{\sin x}{x}$, qui est égal à 1 pour $x = 0$, va en décroissant quand x augmente; par conséquent, il existe une valeur de x, et une seule, pour laquelle $\frac{\sin x}{x} = \frac{1}{2}$. Cette racine de l'équation (18) est évidemment comprise entre $\frac{\pi}{2}$ et $\frac{3\pi}{4}$, ou, pour parler en degrés, entre 90° et 135°.

Soit $f(x) = \log x - \log 2 - \log \sin x$.

Pour $x = 110°$, il vient

$$f(110°) = \log(\text{arc } 110°) - \log 2 - \log \sin 70°$$
$$= +\ 0{,}0092543.$$

La racine étant, d'après cela, très-peu inférieure à 110°, on essaye 109° qui donne encore un résultat positif $+\ 0{,}0026037$; enfin 108° qui donne un résultat négatif $-\ 0{,}0039351$. La racine tombe entre 108° et 109° : on

(*) Euler, *Introduction à l'Analyse infinitésimale*, p. 312.

peut s'assurer, à l'aide de ces trois substitutions, que *la différence seconde*

$$f(110^\circ)-f(109^\circ)-[f(109^\circ)-f(108^\circ)]$$

est plus petite que 0,00012, ce qui légitime l'emploi de la règle des parties proportionnelles.

Cette règle fournira, en partant de 108°, une suite d'approximations *par défaut*, attendu que la dérivée seconde

$$f''(x)=\mathrm{M}\left(\frac{x^2-\sin^2 x}{x^2\sin^2 x}\right)$$

est essentiellement positive.

On trouvera finalement,

$$\text{arc } \mathrm{A}m\mathrm{B} = 108^\circ\, 36'\, 13'',7.$$

100. Problème. — *D'un point* A, *pris sur une circonférence de cercle, tirer deux cordes* AB, AC, *qui divisent l'aire du cercle en trois parties égales* (Euler) (*fig.* 19).

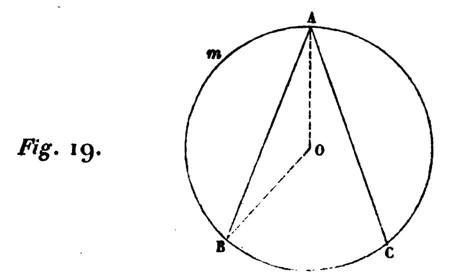


Fig. 19.

En désignant par s l'arc AmB, on aura l'équation

$$s-\sin s=\frac{2}{3}\pi,$$

ou bien,

(19) $$x-\sin(60^\circ-x)=0,$$

en posant $s=x+\frac{2}{3}\pi$.

L'hypothèse $x=\frac{\pi}{6}$, ou 30°, donne dans l'équation (19)

un résultat positif qui ne diffère de zéro que de 0,02 environ. Ainsi la racine doit être très-peu *inférieure* à 30°. En essayant 29°, on trouve un résultat négatif, etc. La méthode des parties proportionnelles, ou celle de Newton, fournira $x = 29^\circ\, 16'\, 27''$, et, par suite,

$$\text{arc } \mathrm{A}m\mathrm{B} = 149^\circ\, 16'\, 27''.$$

101. Problème. — *Déterminer sur une hyperbole équilatère, dont le centre est* O *et le demi-axe transverse* OA, *un point* m *tel, que l'aire* OmA *soit une fraction donnée du triangle* OmP, *qu'on obtient en projetant le point* m *sur l'axe transverse* Ox (*fig.* 20).

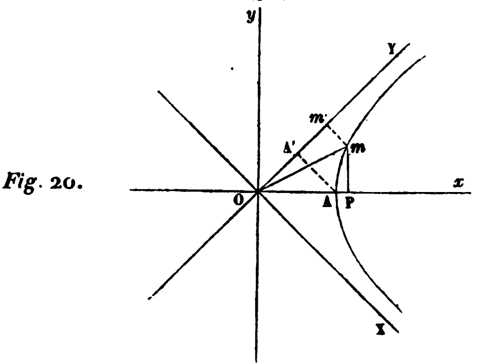


Fig. 20.

Soient OX, OY les asymptotes, et projetons les points A et m sur l'une d'elles en A' et m'. On sait que les deux triangles OAA', Omm' seront équivalents comme moitiés de parallélogrammes équivalents, et, par suite,

$$\text{aire hyperbolique } \mathrm{O}m\mathrm{A} = \text{segment } \mathrm{AA'}mm'.$$

Ce dernier segment est facile à évaluer. Prenons pour unité le demi-axe transverse OA : l'équation de l'hyperbole rapportée à ses asymptotes sera

$$\mathrm{XY} = \frac{1}{2};$$

et l'on aura

$$\text{segment } \mathrm{AA'}mm' = \frac{1}{2}\, l\left(\frac{\mathrm{O}m'}{\mathrm{OA'}}\right)$$

(la caractéristique l désignant un logarithme népérien).

Or $OA' = \frac{1}{\sqrt{2}}$, $Om' = Y$; donc

$$\text{segment } AA'mm' \quad \text{ou} \quad OmA = \frac{1}{2}\, l\,(Y\sqrt{2}).$$

Il reste à revenir de l'ordonnée Y aux coordonnées (x, y) du point m rapportées aux axes principaux de l'hyperbole. Cette transformation donne $Y = \frac{x+y}{\sqrt{2}}$.

Donc

$$\text{aire } OmA = \frac{1}{2}\, l\,(x+y),$$

d'ailleurs,

$$\text{triangle } OmP = \frac{xy}{2}.$$

Si l'on désigne par k le rapport donné de la première aire à la seconde, on aura donc

(20) $$l(x+y) = kxy.$$

On a de plus, entre x et y, l'équation de l'hyperbole

(21) $$x^2 - y^2 = 1.$$

Tel est le système qu'il s'agit de résoudre.

Posons

$$l(x+y) = \frac{u}{2},$$

d'où

$$x+y = e^{\frac{u}{2}},$$

et en vertu de l'équation (21)

$$x-y = e^{-\frac{u}{2}};$$

il en résulte

(22) $$x = \frac{e^{\frac{u}{2}} + e^{-\frac{u}{2}}}{2}, \quad y = \frac{e^{\frac{u}{2}} c - {}^{-\frac{u}{2}}}{2},$$

et substituant ces valeurs dans l'équation (20), on n'aura plus qu'une seule inconnue u. Il vient

$$(23) \qquad 2u = k\,(e^{u} - e^{-u}).$$

Cette équation admet la solution $u = 0$, qui doit être rejetée au point de vue géométrique, puisqu'alors les aires OmA, OmP s'évanouissent.

Si l'on développe le second membre en série, on a

$$e^{u} - e^{-u} = 2u\left(1 + \frac{u^2}{2.3} + \frac{u^4}{2.3.4.5} + \ldots\right),$$

et l'équation (23) se réduit (en supprimant le facteur $2u$) à

$$(24) \qquad \frac{1}{k} - 1 = \frac{u^2}{2.3} + \frac{u^4}{2.3.4.5} + \ldots$$

Cette forme nous montre d'abord que l'équation n'admet que deux racines réelles, égales et de signes contraires. En effet, k étant supposé plus petit que 1, $\left(\frac{1}{k} - 1\right)$ est positif; or le second membre croît avec u d'une manière continue depuis zéro jusqu'à l'infini. Il y a donc une valeur positive de u, et une seule, pour laquelle ce second membre atteindra $\left(\frac{1}{k} - 1\right)$. D'ailleurs la même valeur de u, prise négativement, vérifiera aussi l'équation, puisqu'elle ne renferme que des puissances paires de l'inconnue. Mais il n'en résultera pas deux solutions *géométriques* distinctes du problème. Car lorsqu'on change u en $-u$ dans les formules (22), on voit que x reste le même et y ne fait que changer de signe, en sorte que le point m se transporte symétriquement de l'autre côté de l'axe transverse.

On ne saurait aller plus loin sans assigner une valeur numérique à k. Soit $k = \frac{3}{4}$.

L'équation (24) devient

$$\frac{1}{3}=\frac{u^2}{2.3}+\frac{u^4}{2.3.4.5}+\frac{u^6}{2.3.4.5.6.7}+\dots$$

Cherchons deux limites assez rapprochées entre lesquelles la racine soit comprise. Si d'abord on néglige dans le second membre toutes les puissances de u supérieures à la quatrième, et qu'on résolve l'équation bicarrée

$$\frac{1}{3}=\frac{u^2}{2.3}+\frac{u^4}{2.3.4.5},$$

ou

$$u^4+20u^2-40=0,$$

la valeur positive de u qu'on en tirera sera évidemment plus grande que la racine cherchée : cette valeur est

$$\sqrt{-10+\sqrt{140}}=1,36 \quad \text{(par excès)}.$$

D'autre part, on a

$$\frac{1}{3}<\frac{u^2}{2.3}\left(1+\frac{u^2}{4.5}+\frac{u^4}{(4.5)^2}+\frac{u^6}{(4.5)^3}+\dots\right)$$

ou bien, en sommant cette progression géométrique décroissante,

$$\frac{1}{3}<\frac{u^2}{2.3}\cdot\frac{20}{20-u^2},$$

d'où l'on tire

$$u^2>\frac{20}{11},\ u>\sqrt{\frac{20}{11}}=1,34 \text{ (par défaut)}.$$

Ainsi la racine est comprise entre 1,34 et 1,36; et il va suffire de substituer 1,35 dans l'équation, pour en connaître la valeur à $\frac{1}{100}$ près.

Pour cette substitution, il vaut mieux revenir à la forme (23), parce que les logarithmes s'y appliquent aisément.

En mettant pour k sa valeur $\frac{3}{4}$, on a l'équation

$$3(e^u - e^{-u}) - 8u = 0.$$

Soit

$$f(u) = 3(e^u - e^{-u}) - 8u;$$

l'hypothèse $u = 1,35$ donne successivement :

$$\log e^{1,35} = 1,35 \times 0,43429448 = 0,5862975,$$
$$e^{1,35} = 3,857425,$$
$$\log e^{-1,35} = \bar{1} + (1 - \log e^{1,35}) = \bar{1},4137025,$$
$$e^{-1,35} = 0,259240,$$
$$e^{1,35} - e^{-1,35} = 3,598185,$$
$$3(e^{1,35} - e^{-1,35}) = 10,794555,$$
$$1,35 \times 8 = 10,8,$$
$$f(1,35 = -0,005445.$$

Le signe de ce résultat montre que 1,35 est une valeur approchée, par défaut, à moins de $\frac{1}{100}$.

Pour approcher davantage, on emploiera la méthode de Newton. Comme on a

$$f'(u) = 3(e^u + e^{-u}) - 8,$$

et

$$f''(u) = 3(e^u - e^{-u}),$$

la valeur $u = 1,35$ donnera des résultats de signes contraires pour $f(u)$ et $f''(u)$. Par conséquent (78), en appliquant la méthode à cette valeur, le terme de correction

$$\frac{-f(1,35)}{f'(1,35)}$$

fournira une première approximation par excès.

On trouve

$$f'(1,35) = 4,349995,.$$

d'où

$$-\frac{f(1,35)}{f'(1,35)}=\frac{5445}{4349995}=0,00125.$$

Ce qui nous conduit à pousser ce quotient jusqu'aux cent millièmes, c'est la considération de la limite d'erreur $\frac{M}{2N}h^2$ établie au n° 87. Ici $\frac{M}{2N}$ ou $\frac{f''(1,36)}{2f'(1,35)}$ est visiblement moindre que 2; d'ailleurs h est moindre que 0,002. Donc on a $\frac{M}{2N}h^2<0,000008$.

Ajoutant 0,00125 à 1,35, on a 1,35125, valeur approchée par excès à moins de 0,00001.

Si l'on s'arrête à cette approximation, il restera à substituer cette valeur de u dans les formules (22) qui donneront

$$x=1,2370,\quad y=0,7282.$$

Autre mode de calcul. — Après avoir établi, comme ci-dessus, les équations (20) et (21), posons

$$x+y=\sqrt{z}.$$

L'équation (21) donnera

$$x-y=\frac{1}{\sqrt{z}};$$

d'où

$$(25)\qquad x=\frac{\sqrt{z}+\frac{1}{\sqrt{z}}}{2},\quad y=\frac{\sqrt{z}-\frac{1}{\sqrt{z}}}{2};$$

et, en substituant ces valeurs dans l'équation (20), il viendra

$$\frac{k}{2}\left(z-\frac{1}{z}\right)-lz=0.$$

Soit $k=\frac{3}{4}$. L'équation à résoudre sera

$$(26)\qquad \frac{3}{8}\left(z-\frac{1}{z}\right)-lz=0.$$

Rejetant la solution $z = 1$, pour laquelle les aires s'annulent, on essaye successivement les nombres entiers $z = 2$, $z=3$, $z=4$,... (en prenant les logarithmes de ces nombres dans la Table des logarithmes hyperboliques).

Soit

$$f(z) = \frac{3}{8}\left(z - \frac{1}{z}\right) - lz.$$

Il vient, pour $z = 3$,

$$f(3) = 1 - 1{,}0986\ldots < 0;$$

pour $z = 4$,

$$f(4) = 1{,}406\ldots - 1{,}386\ldots > 0.$$

Le signe ayant changé, il y a une racine comprise entre 3 et 4; et la dérivée $f'(z)$, qui garde constamment le signe $+$ pour toute valeur de z supérieure à 3, nous apprend que cette racine est unique.

Comme la racine paraît plus rapprochée de 4 que de 3, on essayera des nombres équidistants de $\frac{1}{10}$ à partir de 3,9, jusqu'à ce qu'on rencontre un changement de signe.

Voici les calculs.

Pour $z = 3{,}9$,

$$f(3{,}9) = \frac{3}{8}\left(3{,}9 - \frac{1}{3{,}9}\right) - l(3{,}9),$$

$$\frac{3}{8}\left(3{,}9 - \frac{1}{3{,}9}\right) = 1{,}366346;$$

$$l(3{,}9) = l39 - l10 = 3{,}6635616 - 2{,}3025851$$
$$= 1{,}3609765;$$

$$f(3{,}9) = +\,0{,}005370.$$

Pour $z = 3{,}8$, on trouvera de même

$$f(3{,}8) = -\,0{,}008686.$$

Ainsi la racine est comprise entre 3,8 et 3,9.

La méthode de Newton s'appliquera très-bien à la valeur 3,9, pour laquelle $f(z)$ et $f''(z)$ sont de même signe. Nous laisserons au lecteur le soin de développer ces calculs, plus simples que par le précédent mode.

102. Dans l'étude des propriétés de la *chaînette*, on rencontre l'équation suivante :

$$\frac{e^x + e^{-x}}{x} = 2a, \tag{27}$$

a désignant un nombre donné. Il est aisé, d'abord, de reconnaître qu'il existe une valeur *minima* de a, au-dessous de laquelle l'équation est impossible ; car le premier membre devient infini pour $x = 0$ et pour $x = \infty$, tandis qu'il reste fini et positif dans l'intervalle. Lors donc qu'on fait croître x d'une manière continue de zéro à l'infini, ce premier membre atteint nécessairement une certaine valeur plus petite que toutes les autres et supérieure à zéro ; par conséquent, a ne saurait descendre au-dessous de la moitié de cette valeur.

Les considérations géométriques conduisent au même résultat, et, de plus, elles vont nous fournir simplement l'équation propre au calcul du minimum de a, et le nombre des solutions dont l'équation (27) est susceptible quand a dépasse ce minimum. Si l'on pose

$$y = ax \quad \text{et} \quad y = \frac{e^x + e^{-x}}{2},$$

les points d'intersection des lieux géométriques que ces équations représentent auront pour abscisses les racines de l'équation (27). Or l'équation

$$y = \frac{e^x + e^{-x}}{2}$$

représente une chaînette (*fig.* 21), courbe convexe dans tout son cours vers l'axe des x (*), et l'autre équation

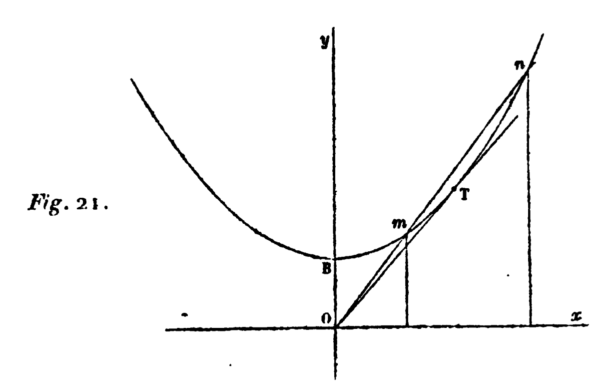


Fig. 21.

définit une droite passant par l'origine O, qui rencontrera la courbe *en deux points* m, n, pourvu que le coefficient angulaire a surpasse celui de la tangente OT. Or on a

$$\operatorname{tang} \mathrm{TO}x = \frac{e^x - e^{-x}}{2} = \frac{e^x + e^{-x}}{2x};$$

La valeur *minima* de a, au-dessous de laquelle le problème est impossible, satisfait donc à la fois aux deux équations

$$a = \frac{e^x - e^{-x}}{2}, \quad a = \frac{e^x + e^{-x}}{2x},$$

et, par conséquent, elle résultera de l'élimination de x entre ces deux équations.

(*) On la construit aisément par points à l'aide des deux *logarithmiques* $y_1 = e^x$, $y_2 = e^{-x}$; car on a $y = \frac{y_1 + y_2}{2}$. Si donc on a construit ces deux logarithmiques (dont la seconde n'est autre que la première transportée symétriquement par rapport à l'axe des y), et que l'on trace, de l'une à l'autre courbe, une série de cordes parallèles à l'axe des y, les points milieux de ces cordes appartiendront à la chaînette.

Posons $e^x = z$, il vient

$$z - \frac{1}{z} = 2a \quad \text{et} \quad z + \frac{1}{z} = 2alz.$$

La première équation donne

$$z = a + \sqrt{a^2+1} \quad \text{(en rejetant la racine négative)},$$

et substituant cette valeur dans la seconde, on a simplement

$$al\left(a + \sqrt{a^2+1}\right) = \sqrt{a^2+1}.$$

Pour résoudre cette équation plus commodément, nous changerons le logarithme népérien en logarithme vulgaire, en multipliant ses deux membres par le module

$$M = 0,43429448\ldots.$$

Elle prendra la forme

$$(28) \qquad a \log\left(a + \sqrt{a^2+1}\right) - M\sqrt{a^2+1} = 0.$$

Soit

$$f(a) = a \log\left(a + \sqrt{a^2+1}\right) - M\sqrt{a^2+1}.$$

L'hypothèse $a = 1$ donne

$$\begin{aligned} f(1) &= \log(1+\sqrt{2}) - M\sqrt{2} \\ &= \log(2,414\ldots) - 0,434\ldots \times 1,414\ldots \\ &= 0,382\ldots - 0,614\ldots \\ &= -0,232\ldots; \end{aligned}$$

$a = 2$ donne

$$\begin{aligned} f(2) &= 2\log(2+\sqrt{5}) - M\sqrt{5} \\ &= 2\log(4,236\ldots) - 0,434 \times 2,236\ldots \\ &= 1,253\ldots - 0,971\ldots \\ &= 0,282\ldots. \end{aligned}$$

Il y a donc une racine comprise entre 1 et 2. D'ailleurs

l'équation (28) n'en admet pas d'autre réelle; car la dérivée

$$f'(a) = \log\left(a + \sqrt{a^2+1}\right)$$

est essentiellement positive, et, par conséquent, le premier membre de l'équation est une fonction continuellement croissante. D'après les résultats de ces deux substitutions, il n'y a pas de raison pour regarder la racine comme plus voisine de 1 que de 2. C'est pourquoi, afin de rapprocher ces limites, nous essayerons 1,5 :

$$f(1,5) = (1,5)\log\left(1,5 + \sqrt{3,25}\right) - \mathrm{M}\sqrt{3,25},$$

$\log(3,25) = 0,5118834$	
$\log\sqrt{3,25} = 0,2559417$	$\log\sqrt{3,25} = 0,2559417$
$\sqrt{3,25} = 1,802775$	$\log \mathrm{M} = \bar{1},6377843$
$1,5 + \sqrt{3,25} = 3,302775$	
$\log\left(1,5 + \sqrt{3,25}\right) = 0,5188790$	$\log\left(\mathrm{M}\sqrt{3,25}\right) = \bar{1},8937260$
$\log\left(1,5 + \sqrt{3,25}\right) = 0,7783185$	$\mathrm{M}\sqrt{3,25} = 0,7829355$

$$f(1,5) = 0,7783185 - 0,7829355 = -0,0046170.$$

La racine surpasse donc 1,5, mais elle en est très-rapprochée. En essayant 1,6, on trouve

$$f(1,6) = +0,0484574.$$

Ainsi la racine est comprise entre 1,5 et 1,6.

On appliquera maintenant la méthode de Newton. Bien que $f(a)$ et $f''(a)$, dont l'expression est $\frac{\mathrm{M}}{\sqrt{a^2+1}}$, soient de signes contraires pour $a = 1,5$, c'est de 1,5 qu'il convient de partir, attendu que ce nombre est beaucoup plus approché que 1,6. Le terme de correction

$$-\frac{f(1,5)}{f'(1,5)} = \frac{4617}{518879}$$

fournira une approximation par excès; ce quotient sera évidemment moindre que 0,01. Donc la limite d'erreur qu'il comporte, et qui a pour expression (87)

$$\frac{f''(a)}{2f'(a)}h^2 = \frac{\mathrm{M}}{2\sqrt{a^2+1}\log(a+\sqrt{a^2+1})}h^2,$$

est moindre que

$$\frac{0,5}{2.1,8.0,5}\,0,0001 < 0,00003;$$

et, par conséquent, on doit pousser la division jusqu'aux dix millièmes : il vient

$$\frac{4617}{518879} = 0,0089 \quad \text{(par excès)};$$

d'où l'on conclut

$$a_1 = 1,5089;$$

telle est la valeur du minimum de a, approchée par excès à moins de $\frac{1}{10000}$.

Le calcul d'un second terme de correction $-\frac{f(a_1)}{f'(a_1)}$, fournirait certainement huit décimales exactes.

Toutes les fois que a surpassera le *minimum* 1,5089, il résulte des considérations géométriques exposées plus haut que l'équation

$$(27) \qquad \frac{e^x + e^{-x}}{x} = 2a$$

admettra *deux* valeurs réelles pour x. Ces deux solutions se réduiront à une seule, quand a sera égal à son minimum.

Par exemple, si l'on fait $a = 2$, l'équation

$$e^x + e^{-x} - 4x = 0$$

aura deux racines réelles, l'une comprise entre 0 et 1, l'autre entre 2 et 3. Nous laissons au lecteur le soin de les calculer approximativement.

103. Bien que la méthode de Newton ait été appliquée, dans l'exemple qui précède, avec grand succès, il sera bon d'indiquer ici une *méthode d'approximations successives par substitution*, que nous n'avons pas encore eu l'occasion d'employer, et dont la théorie est des plus simples.

Cette méthode consiste à négliger d'abord dans l'équation, préparée sous forme convenable, une certaine classe de termes, de manière à tirer de l'équation ainsi tronquée une première valeur approchée de la racine. Puis on substitue cette valeur à la place de l'inconnue dans l'ensemble des termes négligés, ce qui les réduit à un nombre, dont on tient compte pour corriger la première approximation. Il en résulte une seconde valeur approchée avec laquelle on en calcule de même une troisième, et ainsi de suite. Souvent il arrive qu'on obtient ainsi une suite de nombres, *alternativement* plus petits et plus grands que la racine cherchée; ils fournissent alors une approximation dont le sens et le degré sont parfaitement définis à toutes les époques de l'opération.

Reprenons l'équation

$$a\,l\left(a+\sqrt{a^2+1}\right)=\sqrt{a^2+1},$$

et mettons-la d'abord sous la forme

$$a+\sqrt{a^2+1}=e^{\sqrt{1+\frac{1}{a^2}}};$$

si l'on multiplie les deux membres par $a-\sqrt{a^2+1}$, on en tire

$$a-\sqrt{a^2+1}=-e^{-\sqrt{1+\frac{1}{a^2}}};$$

ajoutant ces deux équations membre à membre, on élimine

le radical du premier membre, et il vient

$$a = \frac{e^{\sqrt{1+\frac{1}{a^2}}} - e^{-\sqrt{1+\frac{1}{a^2}}}}{2}.$$

Enfin, on développe le second membre en série, et l'on a

$$(29)\quad a = \sqrt{1+\frac{1}{a^2}}\left[1+\frac{1+\frac{1}{a^2}}{1.2.3}+\frac{\left(1+\frac{1}{a^2}\right)^2}{1.2.3.4.5}+\frac{\left(1+\frac{1}{a^2}\right)^3}{1.2.3.4.5.6.7}+\ldots\right].$$

Telle est la forme la plus propre aux substitutions successives que nous allons faire.

1°. On néglige $\frac{1}{a^2}$ dans le second membre qui devient ainsi trop petit, et, par suite, on a

$$a > 1+\frac{1}{1.2.3}+\frac{1}{1.2.3.4.5}+\frac{1}{1.2.3.4.5.6.7}+\ldots.$$

En additionnant quelques termes de cette série, les trois premiers, par exemple, on trouve

$$a > 1,17.$$

Posons $a_1 = 1,17$. Puisque a_1 est trop petit, $\frac{1}{a_1}$ est trop grand. Si donc on remplace $\frac{1}{a^2}$ par $\frac{1}{a_1^2}$ dans le second membre, il viendra

$$a < \sqrt{1+\frac{1}{(1,17)^2}}\left[1+\frac{1+\frac{1}{(1,17)^2}}{1.2.3}+\frac{\left(1+\frac{1}{(1,17)^2}\right)^2}{1.2.3.4.5}+\ldots\right].$$

Le calcul par excès des termes écrits ci-dessus suffit pour donner

$$a < 1,73.$$

Soit $a_2 = 1,73$. De $a < a_2$, il résulte $\frac{1}{a} > \frac{1}{a_2}$, et, par

suite, en substituant $\frac{1}{a_2}$ dans le second membre de l'équation (29), on aura

$$a > \sqrt{1 + \frac{1}{(1,73)^2}} \left[1 + \frac{1 + \frac{1}{(1,73)^2}}{1.2.3} + \frac{\left(1 + \frac{1}{(1,73)^2}\right)^2}{1.2.3.4.5} + \ldots \right].$$

Soit a_3 une valeur par défaut du second membre, on aura

$$a > a_3,$$

et l'on continuera ainsi jusqu'à ce qu'on ait une approximation suffisante.

Les nombres a_1, a_3, a_5,..., tous plus petits que la racine, forment une suite évidemment croissante, tandis que les nombres a_2, a_4, a_6, tous plus grands que la racine, forment une suite décroissante : par conséquent, les deux suites sont convergentes vers cette même racine. Lorsqu'on s'arrête à un terme quelconque, l'erreur commise est moindre que la différence entre ce terme et celui qui le suit dans l'autre série ; cette différence va constamment en diminuant à mesure que le rang du terme est plus élevé. Toutefois, l'approximation marche avec un peu de lenteur dans l'exemple ci-dessus : cela tient à ce que a étant peu supérieur à 1, les puissances de $\frac{1}{a^2}$ ne peuvent être négligées sans qu'il en résulte une erreur notable.

104. Nous proposerons, comme application de la même méthode, la question suivante :

Une somme de 3000 *francs, placée à intérêts composés pendant* 14 *ans* 2 *mois et* 12 *jours, est devenue au bout de ce temps* 5605fr,84. *On demande le taux d'intérêt.*

(Il est convenu qu'on ne capitalise les intérêts que pendant les années, et la somme résultante est ensuite placée à

intérêts simples pendant les mois et jours qui suivent. — On fait tous les mois égaux à 30 jours.)

En désignant par r l'intérêt d'un franc pour un an, l'équation du problème est

$$3000(1+r)^{14}\left(1+\frac{72\,r}{360}\right)=5605,84\,;$$

d'où l'on tire

$$\log(1+r)=\frac{\log(5605,84)-\log 3000}{14}-\frac{\log\left(1+\frac{r}{5}\right)}{14}.$$

Si d'abord on néglige le terme $\dfrac{\log\left(1+\frac{r}{5}\right)}{14}$ dont la valeur est petite comparée à $\log(1+r)$, on aura

$$\log(1+r)<\frac{\log 5605,84-\log 3000}{14},$$
$$<0,0193943\,;$$

d'où

$$1+r<1,046.$$

Soit $r_1=0,046$. r_1 est une première valeur approchée par excès du taux d'intérêt cherché. On remplace maintenant r par r_1 dans le terme négligé et l'on a

$$\log(1+r)>0,0193942-\frac{\log\left(1+\frac{0,046}{5}\right)}{14}$$
$$>0,0191101\,;$$

d'où

$$1+r>1,0449.$$

Soit $r_2=0,0449$. r_2 est une valeur approchée par défaut du taux d'intérêt.

En substituant r_2 à la place de r dans le terme $\dfrac{\log\left(1+\frac{r}{5}\right)}{14}$, on trouvera de même

$$r_3=0,045002,$$

valeur par excès, et ainsi de suite. La somme proposée était donc placée, à très-peu près, à $4\frac{1}{2}$ pour 100.

105. Cette méthode d'approximation par substitutions successives s'emploie encore avec avantage pour résoudre les équations du second et du troisième degré, dans les cas où le coefficient de la plus haute puissance de l'inconnue est un nombre très-petit. C'est par là que nous terminerons cette théorie des approximations numériques.

Second degré : $ax^2 \pm bx \pm c = 0$.

a désigne un nombre très-petit que l'on peut toujours supposer positif; en ayant égard aux signes de b et c, on aura quatre cas à distinguer, dont deux seulement demandent à être traités séparément,

$$ax^2 + bx - c = 0 \quad , \quad ax^2 - bx + c = 0;$$

car les deux autres se ramènent à ceux-ci par le changement de x en $-x$.

1°. $ax^2 + bx - c = 0$.

On sait, par la discussion de la formule $\frac{-b \pm \sqrt{b^2 + 4ac}}{2a}$, que l'une des racines tend, pour des valeurs décroissantes de a, vers la limite $\frac{c}{b}$, tandis que l'autre croît indéfiniment. C'est de la plus petite racine, celle qui diffère peu de $\frac{c}{b}$, que nous allons nous occuper : l'autre s'en déduira par différence, puisque la somme des deux racines est égale à $\frac{-b}{a}$.

Nous mettrons l'équation sous la forme

$$x = \frac{c}{b} - \frac{a}{b}x^2.$$

Si d'abord on néglige le terme $\frac{a}{b}x^2$, qui est de l'ordre

de a, on a une *première approximation par excès*,

$$x_1 = \frac{c}{b}.$$

Remplaçant x par x_1 dans le terme négligé, on retranchera de $\frac{c}{b}$ une quantité $\frac{a}{b}x_1^2$ trop forte, d'où résultera une *seconde approximation par défaut*

$$x_2 = \frac{c}{b} - \frac{ac^2}{b^3}.$$

L'erreur commise en prenant $x_1 = \frac{c}{b}$ est donc moindre que la différence $\frac{ac^2}{b^3}$; elle est au plus de l'ordre de a, comme nous l'avions dit.

La valeur x_2 n'est en défaut que d'une quantité de l'ordre de a^2, qu'on appréciera plus bas. Substituant x_2 dans le second membre de l'équation, on retranchera de $\frac{c}{b}$ une quantité $\frac{a}{b}x_2^2$ trop faible. De là une *troisième approximation par excès*

$$\frac{c}{b} - \frac{a}{b}\left(\frac{c}{b} - \frac{ac^2}{b^3}\right)^2.$$

Mais, en développant le carré, on trouverait un terme $\frac{a^3c^4}{b^7}$, de l'ordre de a^3, qu'il serait illusoire de conserver. En effet, x_2 est en défaut d'une quantité inconnue α qui est de l'ordre de a^2; l'erreur qui affecte $\frac{a}{b}x_2^2$, c'est-à-dire

$$\frac{a}{b}\left[(x_2 + \alpha)^2 - x_2^2\right] = \frac{2a\alpha x_2}{b} + \ldots,$$

est donc déjà de l'ordre de a^3. Par conséquent, on ne doit

conserver dans le développement du carré $\left(\frac{c}{b}-\frac{ac^2}{b^3}\right)^2$ que les termes du premier et du second ordre, et l'on adoptera, comme troisième approximation,

$$x_3=\frac{c}{b}-\frac{ac^2}{b^3}+\frac{2a^2c^3}{b^5}.$$

Cette valeur sera, à fortiori, approchée par excès, puisqu'on a négligé un terme négatif. Si on la compare à x_2, on voit que l'erreur commise en prenant x_2 à la place de x, erreur que nous désignions plus haut par α, est moindre que $\frac{2a^2c^3}{b^5}$; elle est donc de l'ordre de a^2.

En continuant ainsi, on trouvera l'expression

$$\frac{c}{b}-\frac{a}{b}\left(\frac{c}{b}-\frac{ac^2}{b^3}+\frac{2a^2c^3}{b^5}\right)^2$$

dont il ne faudra conserver, pour la raison susdite, que les termes qui ne dépassent pas le troisième ordre, c'est-à-dire $\frac{c}{b}-\frac{a}{b}\left[\left(\frac{c}{b}-\frac{ac^2}{b^3}\right)^2+\frac{4a^2c^4}{b^6}\right]$. Ainsi on aura pour *quatrième valeur approchée par défaut*

$$x_4=\frac{c}{b}-\frac{ac^2}{b^3}+\frac{2a^2c^3}{b^5}-\frac{5a^3c^4}{b^7}.$$

(Nous disons qu'elle est *par défaut*, bien qu'on ait négligé plusieurs termes : car le premier des termes négligés $\frac{4a^4c^5}{b^9}$ est positif et du quatrième ordre, tandis que le suivant $\frac{-4a^5c^6}{b^{11}}$ est du cinquième, et, par suite, beaucoup plus petit. On néglige donc en définitive une quantité positive dans une expression qui était déjà par défaut.)

Maintenant, on a une limite supérieure de l'erreur commise sur x_3 : cette erreur est moindre que $\frac{5a^3c^4}{b^7}$, etc.

On aperçoit bien que les termes dont se composent les valeurs approchées x_2, x_3, x_4, etc., sont alternativement positifs et négatifs; que les exposants de b croissent comme les nombres impairs, tandis que ceux de a et de c croissent comme la suite naturelle des nombres. Mais la loi des coefficients ne se manifeste pas.

Cette loi serait mise en évidence par le développement en série convergente de la racine $\frac{-b+\sqrt{b^2+4ac}}{2a}$, qu'on écrirait d'abord sous la forme $\frac{b}{2a}\left[-1+\left(1+\frac{4ac}{b^2}\right)^{\frac{1}{2}}\right]$. Puis appliquant la formule du binôme de Newton (supposée démontrée pour un exposant quelconque), on aurait la série convergente

$$\frac{b}{2a}\left[\frac{1}{2}\cdot\frac{4ac}{b^2}-\frac{1}{2.4}\left(\frac{4ac}{b^2}\right)^2+\frac{1.3}{2.4.6}\left(\frac{4ac}{b^2}\right)^3-\frac{1.3.5}{2.4.6.8}\left(\frac{4ac}{b^2}\right)+\ldots\right].$$

En effectuant les puissances indiquées et réduisant, cette série fournirait, non-seulement les termes déjà calculés plus haut, mais tous ceux des ordres supérieurs. Au reste, dans la pratique où a est supposé très-petit, on ne pousse guère le calcul au delà de la valeur x_2.

2°. $ax^2-bx+c=0$.

On s'attache, comme dans le cas précédent, à calculer la plus petite racine, celle dont la valeur numérique diffère peu de $\frac{c}{b}$. On met, à cet effet, l'équation sous la forme

$$x=\frac{c}{b}+\frac{a}{b}x^2.$$

Si d'abord on néglige $\frac{a}{b}x^2$, on a pour *première valeur approchée par défaut*,

$$x_1=\frac{c}{b}.$$

Substituant cette valeur à la place de x dans le terme négligé, on ajoute à $\frac{c}{b}$ une quantité trop faible, d'où résulte une *seconde valeur approchée par défaut,* comme la première

$$x_2 = \frac{c}{b} + \frac{ac^2}{b^3}.$$

On n'a plus ici de limite supérieure pour l'erreur commise à chaque approximation. On sait seulement que x_1 est approchée par défaut à moins d'une quantité de l'ordre de a; que x_2 est approchée dans le même sens à moins d'une quantité de l'ordre de a^2. En continuant, on trouvera

$$x_3 = \frac{c}{b} + \frac{ac^2}{b^3} + \frac{2a^2c^3}{b^5},$$

valeur dans laquelle on a négligé, comme dans le premier cas, le terme en a^3 provenant du carré de $\left(\frac{c}{b} + \frac{ac^2}{b^3}\right)$, parce qu'on ne peut répondre de l'exactitude des termes de cet ordre.

Au reste, il suffisait de changer a en $-a$ dans tous les résultats obtenus précédemment. Toutes ces valeurs sont approchées par défaut, et comme chacune est plus grande que la précédente, l'approximation va croissant.

106. *Application.* — Soit l'équation

$$0,00018.x^2 + 4,3974.x - 3,5662 = 0,$$
$$a = 0,00018, \quad b = 4,3974, \quad c = 3,5662.$$

On vient de trouver pour valeur approchée de la plus petite racine

$$x = \frac{c}{b} - \frac{ac^2}{b^3} + \frac{2a^2c^3}{b^5} - \ldots.$$

Si l'on a recours aux logarithmes, on trouvera

$$\log \frac{c}{b} = \bar{1},9090097,$$

d'où

$$\frac{c}{b} = 0,8109790.$$

La première valeur approchée par excès est donc

$$x_1 = 0,8109790.$$

Puis

$$\log \frac{ac^2}{b^3} = \bar{5},4300959,$$

d'où

$$\frac{ac^2}{b^3} = 0,0000269.$$

Nous ne conservons que trois chiffres significatifs (ou sept décimales) pour $\frac{ac^2}{b^3}$, bien que les Tables de logarithmes en donnent encore deux autres; parce que ce terme doit être retranché de $\frac{c}{b}$ dont la septième décimale 0 est déjà douteuse. Il vient ainsi, pour seconde valeur approchée par défaut,

$$x_2 = 0,8109521.$$

Les Tables ordinaires ne permettent pas de pousser l'approximation plus loin; car le terme suivant, $\frac{2a^2c^3}{b^5}$, aurait pour caractéristique de son logarithme $\bar{9}$, et, par suite, il n'influerait pas sur les sept premières décimales. Toutefois, comme ce terme est une limite supérieure de l'erreur qui peut affecter x_2, il en résulte que cette valeur est approchée à moins d'une unité de son dernier ordre décimal.

2°. On sait que la profondeur d'un puits peut être déter minée en y laissant tomber une pierre et observant le temps qui s'écoule depuis le moment où on lâche la pierre jusqu'à celui où le bruit qu'elle fait en frappant l'eau par-

vient à l'oreille. Soit t ce temps exprimé en secondes. On emprunte à l'expérience deux autres données : la vitesse v du son, ou le nombre de mètres que le son parcourt par seconde ($v = 340^{m}$); et l'espace $\frac{g}{2}$ parcouru par un corps pesant dans la première seconde de chute ($g = 9^{m},8088$). En désignant par x la profondeur du puits, on aura $\frac{x}{v}$ pour l'expression du temps que met le son à monter uniformément jusqu'à l'oreille, et $\sqrt{\frac{2x}{g}}$ pour le temps que met la pierre à descendre d'un mouvement uniformément accéléré jusqu'à l'eau du puits. La somme de ces deux temps doit être égale à t : l'équation du problème est donc

$$\frac{x}{v} + \sqrt{\frac{2x}{g}} = t. \tag{30}$$

En isolant le radical, puis élevant au carré, il vient

$$\frac{x^2}{v^2} - 2\left(\frac{v}{g} + t\right)\frac{x}{v} + t^2 = 0. \tag{31}$$

Cette équation a ses deux racines réelles et positives; mais il faut remarquer qu'une seule de ces racines satisfera à la question, c'est-à-dire à l'équation (30) dans laquelle le radical est pris avec le signe $+$. Si l'on trouve deux racines, c'est que l'équation (31) comprend en outre les solutions de l'équation

$$\frac{x}{v} - \sqrt{\frac{2x}{g}} = t,$$

qui ne diffère de (30) que par le signe du radical. Pour distinguer laquelle des racines de l'équation (31) convient au problème, il suffit d'observer que le produit des deux valeurs de $\frac{x}{v}$ étant égal à t^2, l'une d'elles est plus petite que t et l'autre plus grande. Or, d'après l'équa-

tion (30), $\frac{x}{v}$ est évidemment plus petit que t. Donc c'est la plus petite racine de l'équation (31) qui doit être adoptée.

Le coefficient $\frac{1}{v^2}$ de x^2 est moindre que $\frac{1}{100000}$, tandis que le coefficient de x et le terme tout connu sont des nombres bien supérieurs. Nous pouvons donc appliquer au calcul de la plus petite racine la formule d'approximation du second cas :

$$x = \frac{c}{b} + \frac{ac^2}{b^3} + \ldots.$$

Nous avons ici

$$a = \frac{1}{v^2} = \frac{1}{(340)^2}, \quad b = 2\left(\frac{1}{g} + \frac{t}{v}\right), \quad c = t^2.$$

Supposons, par exemple, $t = 4^{sec},35$. On trouvera

$$\log \frac{c}{b} = 1,9162209, \quad \text{d'où} \quad \frac{c}{b} = 82,455....$$

Une première valeur approchée par défaut est donc

$$x_1 = 82^m,4.$$

Puis on a

$$\log \frac{ac^2}{b^3} = \bar{1},4087263, \quad \text{d'où} \quad \frac{ac^2}{b^3} = 0,2562....$$

Ajoutant ce nombre à $\frac{c}{b}$, on obtient

$$x_2 = 82^m,71,$$

valeur également par défaut, et d'une approximation bien suffisante.

107. *Remarque.* — Si l'on supposait c très-petit, au lieu de a, dans l'équation

$$ax^2 \pm bx \pm c = 0,$$

il ne serait pas nécessaire de recommencer une discussion nouvelle ; car, en posant $x = \frac{1}{y}$, l'équation devient

$$cy^2 \pm by \pm a = 0,$$

et rentre, par conséquent, dans les deux cas déjà traités.

L'une des valeurs de y différera peu de $\frac{a}{b}$ (abstraction faite du signe), et l'autre aura une valeur numérique très-grande. On pourra calculer la première par les formules d'approximation qui viennent d'être établies ; puis prendre le logarithme de cette valeur, changer son signe, et l'on aura le logarithme de la valeur correspondante de x. Cette racine sera voisine de $\frac{b}{a}$. Quant à l'autre, elle sera très-petite et se conclura de la précédente, puisque la somme est égale à $\pm \frac{b}{a}$.

108. *Troisième degré :* $ax^3 \pm bx \pm c = 0$.

On suppose a très-petit et positif. La condition de réalité des racines

$$\pm 4b^3 + 27ac^2 < 0,$$

sera toujours satisfaite, *si* b *est précédé du signe* —. Dans ce cas, si l'on met les signes de c en évidence, on aura les deux équations

$$ax^3 - bx + c = 0, \quad ax^3 - bx - c = 0,$$

dont la seconde se déduit de la première en changeant x en $-x$. Il suffit donc de traiter la première. On en tire

$$x = \frac{c}{b} + \frac{a}{b} x^3.$$

Négligeant dans une première approximation $\frac{a}{b} x^3$, il vient

$$x_1 = \frac{c}{b},$$

valeur approchée par défaut. Puis on aura successivement, comme on l'a vu pour le second degré,

$$x_2 = \frac{c}{b} + \frac{ac^3}{b^4},$$

$$x_3 = \frac{c}{b} + \frac{ac^3}{b^4} + \frac{3a^2c^5}{b^7},$$

valeurs toutes par défaut, et de plus en plus approchées de l'une des racines de l'équation proposée.

Quant aux deux autres, il est aisé de voir qu'elles ont des valeurs numériques très-grandes et de signes contraires. En effet, si l'on pose $x = \frac{1}{y}$, il vient

$$cy^3 - by^2 + a = 0,$$

ou

$$y^2(b - cy) = a.$$

Or, pour rendre le produit $y^2(b - cy)$ très-petit, on peut s'y prendre de trois manières : soit en faisant y très-peu inférieur à $\frac{b}{c}$ $\left(\text{et, par suite, } x \text{ peu supérieur à } \frac{c}{b}, \text{ c'est la racine déjà calculée}\right)$; soit en donnant à y une valeur très-petite, positive ou négative, auquel cas l'autre facteur différera peu de b. A ces deux dernières valeurs de y, correspondront pour x deux valeurs très-grandes et de signes contraires.

On calcule aisément y par approximations successives. Ainsi d'abord, négligeant y^3 dans l'équation

$$cy^3 - by^2 + a = 0,$$

il vient

$$y_1 = \pm\sqrt{\frac{a}{b}}.$$

Prenons le signe $+$, et pour approcher davantage de y, posons

$$y = y_1 + \alpha.$$

En désignant le premier membre de l'équation par $f(y)$, la substitution de $y_1 + \alpha$ donne

$$(32) \qquad f(y_1) + f'(y_1)\alpha + \frac{f''(y_1)}{1.2}\alpha^2 + c\alpha^3 = 0.$$

Comme α est petit, on néglige ses puissances supérieures à la première, et l'on a, approximativement,

$$f(y_1) + f'(y_1)\alpha = 0.$$

D'ailleurs

$$f(y_1) = cy_1^3 - by_1^2 + a = cy_1^3,$$

vu que $(-by_1^2 + a)$ est nul d'après la valeur de y_1 ; et

$$f'(y_1) = 3cy_1^2 - 2by_1.$$

On a donc, pour déterminer α, l'équation

$$(33) \qquad cy_1^3 + (3cy_1 - 2b)\alpha = 0;$$

mais elle doit encore être simplifiée. En effet, il est évident que la valeur de α qu'on en tirera sera du même ordre de grandeur que y_1^2, ou de l'ordre de a, puisque y_1 est de l'ordre de $a^{\frac{1}{2}}$; par suite, le terme $3cy_1\alpha$ est de l'ordre de $a^{\frac{3}{2}}$. Or, en omettant, comme on l'a fait ci-dessus, les termes en α^2 de l'équation (32), on a déjà implicitement négligé une quantité de l'ordre de $\frac{\alpha^2}{y_1}$ ou de $a^{\frac{3}{2}}$. Il serait donc illusoire de conserver dans l'équation (33) le terme $3cy_1\alpha$. En conséquence, on tirera simplement

$$\alpha = \frac{cy_1^2}{2b} = \frac{ac}{2b^2};$$

de là, une seconde valeur approchée,

$$y_2 = \sqrt{\frac{a}{b}} + \frac{ac}{2b^2}.$$

Nota. — Il n'a pas échappé au lecteur que la valeur de α qu'on tire de l'équation $f(y_1) + f'(y_1)\alpha = 0$, est précisément le terme de correction de la méthode de Newton, $-\frac{f(y_1)}{f'(y_1)}$. Mais nous avons dû réduire ce terme à la partie qui est de l'ordre de a.

Maintenant, pour pousser l'approximation jusqu'aux termes de l'ordre de $a^{\frac{3}{2}}$, on posera dans l'équation primitive

$$y = y_2 + \beta,$$

et l'on aura, en changeant y_1 en y_2 et α en β, l'équation approchée

$$f(y_2) + f'(y_2)\beta = 0,$$

ou

$$f\left(y_1 + \frac{cy_1^2}{2b}\right) + f'\left(y_1 + \frac{cy_1^2}{2b}\right)\beta = 0.$$

Comme β sera de l'ordre de y_1^3, on négligera dans $f\left(y_1 + \frac{cy_1^2}{2b}\right)$ les termes d'ordre supérieur à y_1^4, et dans $f'\left(y_1 + \frac{cy_1^2}{2b}\right)$ les termes d'ordre supérieur à y_1. Il vient ainsi

$$f\left(y_1 + \frac{cy_1^2}{2b}\right) = f(y_1) + f'(y_1)\frac{cy_1^2}{2b} + \frac{f''(y_1)}{1.2}\frac{c^2y_1^4}{4b^2}$$
$$= \frac{5c^2y_1^4}{4b},$$
$$f\left(y_1 + \frac{cy_1^2}{2b}\right) = -2by_1;$$

d'où

$$\beta = \frac{5c^2y_1^3}{8b^2} = \frac{5a\sqrt{a}c^2}{8b^3\sqrt{b}}.$$

On a donc pour troisième approximation de y,

$$y_3 = \sqrt{\frac{a}{b}} + \frac{ac}{2b^2} + \frac{5a\sqrt{a}c^2}{8b^3\sqrt{b}}.$$

En s'arrêtant là, calculant le logarithme de cette valeur, puis changeant son signe, on aura le logarithme approché de la valeur correspondante de x.

Si l'on opère de même pour la seconde racine, on a d'abord

$$y_1 = -\sqrt{\frac{a}{b}}.$$

Puis, en posant $y = y_1 + \alpha$, on voit que la valeur approchée de α, où y_1 n'entre qu'au carré, ne différera pas de la précédente

$$\alpha = \frac{ac}{2b^2}.$$

Au contraire, β changera de signe, et deviendra

$$\beta = -\frac{5a\sqrt{ac^2}}{8b^3\sqrt{b}}.$$

La seconde racine de l'équation en y sera donc, approximativement,

$$-\sqrt{\frac{a}{b}}+\frac{ac}{2b^2}-\frac{5a\sqrt{ac^2}}{8b^3\sqrt{b}};$$

on en conclura, comme ci-dessus, la valeur correspondante de x.

Il reste à traiter le cas où *b est précédé du signe* + dans l'équation du troisième degré. En ayant égard au signe de c, on a les deux équations

$$ax^3+bx-c=0, \quad ax^3+bx+c=0.$$

La seconde se déduit de la première en changeant x en $-x$. Il suffit donc de considérer celle-ci. On la met sous la forme

$$x=\frac{c}{b}-\frac{a}{b}x^3.$$

Or cette équation ne diffère de celle que nous avons traitée dans le premier cas, qu'en ce que a est changé en $-a$. On aura donc

$$x_1=\frac{c}{b},$$

$$x_2=\frac{c}{b}-\frac{ac^3}{b^4},$$

$$x_3=\frac{c}{b}-\frac{ac^3}{b^4}+\frac{3a^2c^5}{b^7},$$

valeurs alternativement par excès et par défaut; l'erreur que comporte chacune d'elles est moindre que le terme de correction qui s'ajoute ou se retranche de cette valeur pour former la suivante.

Quant aux deux autres racines, il n'y a pas lieu de les calculer approximativement, puisqu'elles sont imaginaires.

FIN.

